



US009476032B2

(12) **United States Patent**
Wimmer et al.

(10) **Patent No.:** US 9,476,032 B2
(45) **Date of Patent:** Oct. 25, 2016

(54) **ATTENUATED VIRUSES USEFUL FOR VACCINES**

(75) Inventors: **Eckard Wimmer**, East Setauket, NY (US); **Steve Skiena**, Setauket, NY (US); **Steffen Mueller**, Kings Point, NY (US); **Bruce Futcher**, Stony Brook, NY (US); **Dimitris Papamichail**, South Miami, FL (US); **John Robert Coleman**, Blauvelt, NY (US); **Jeronimo Cello**, Port Jefferson, NY (US)

(73) Assignee: **The Research Foundation for The State University of New York**, Albany, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 902 days.

(21) Appl. No.: **12/594,173**

(22) PCT Filed: **Mar. 31, 2008**

(86) PCT No.: **PCT/US2008/058952**

§ 371 (c)(1),
(2), (4) Date: **Mar. 29, 2010**

(87) PCT Pub. No.: **WO2008/121992**

PCT Pub. Date: **Oct. 9, 2008**

(65) **Prior Publication Data**

US 2010/0209454 A1 Aug. 19, 2010

Related U.S. Application Data

(60) Provisional application No. 60/909,389, filed on Mar. 30, 2007, provisional application No. 61/068,666, filed on Mar. 7, 2008.

(51) **Int. Cl.**

C12N 7/04 (2006.01)
A61K 39/00 (2006.01)
C12N 7/00 (2006.01)

(52) **U.S. Cl.**

CPC **C12N 7/00** (2013.01); **A61K 2039/5254** (2013.01); **C12N 2720/12361** (2013.01); **C12N 2740/15061** (2013.01); **C12N 2760/16061** (2013.01); **C12N 2770/20061** (2013.01); **C12N 2770/24161** (2013.01); **C12N 2770/32061** (2013.01); **C12N 2770/32661** (2013.01); **C12N 2770/32761** (2013.01)

(58) **Field of Classification Search**

CPC **A61K 2039/5254**; **C12N 7/00**; **C12N 2720/12361**; **C12N 2740/15061**; **C12N 2770/32761**; **C12N 2770/20061**; **C12N 2770/24161**; **C12N 2770/32061**; **C12N 2770/32661**; **C12N 2760/16061**

USPC 424/229

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,696,289 B1 2004/0097439 A9 2004/0209241 A1 2008/0118530 A1 *	2/2004 Bae et al. 5/2004 Nicholas 10/2004 Hermason et al. 5/2008 Kew et al. 424/207.1
---	---

FOREIGN PATENT DOCUMENTS

WO 02/095363 A2 WO 2006042156	11/2002 * 4/2006
----------------------------------	---------------------

OTHER PUBLICATIONS

- Burns, et al. Modulation of Poliovirus Replicative Fitness in HeLa Cells by Deoptimization of Synonymous Codon Usage in the Capsid Region. *J. Virol.* 2006; 80(7): 3259-3272.*
Mueller et al., Reduction of the Rate of Poliovirus Protein Synthesis through Large-Scale Codon Deoptimization Causes Attenuation of Viral Virulence by lowering Specific Infectivity, 2006, *Journal of Virology*, 80(19): pp. 9687-9696.*
Cheng, L. et al., "Absence of Effect of Varying Thr-Leu Codon Pairs on Protein Synthesis in a T7 System", *Biochem.* (2001), vol. 40, pp. 6102-6106.
Cohen, B. et al., "Natural Selection and Algorithmic Design of mRNA", *B JCB* (2003), vol. 10, pp. 3-4.
Doma, M. et al., "Endonucleolytic Cleave of Eukaryotic mRNAs with Stalls and Translation Elongation", *Nature* (2006), vol. 440, pp. 561-564.
Garcia-Sastre, A. et al., "Genetic Manipulation of Negative-Strand RNA Virus Genomes", *Annu. Rev. Microbiol.* (1993), vol. 47, pp. 765-790.
Greve, J. et al., "The Major Human Rhinovirus Receptor is ICAM-1", *Cell* (1989), vol. 56, pp. 839-847.
Gustafsson, C. et al., "Codon Bias and Heterologous Protein Expression", *Trends in Biotechnology* (2004), vol. 22:7, pp. 346-353.
Johansen, L. et al., "The RNA Encompassing the Internal Ribosome Entry in the Poliovirus 5' Nontranslated Region Enhances the Encapsidation of Genomic RNA", *Virology* (2000), vol. 273, pp. 391-399.
Lavner, Y et al., "Codon Bias as a Factor in Regulating Expression via Translation Rate in the Human Genome", *Gene* (2005), vol. 345, pp. 127-138.

(Continued)

Primary Examiner — Nicole Kinsey White

(74) *Attorney, Agent, or Firm* — Fox Rothschild LLP

(57) **ABSTRACT**

This invention provides an attenuated virus which comprises a modified viral genome containing nucleotide substitutions engineered in multiple locations in the genome, wherein the substitutions introduce synonymous deoptimized codons into the genome. The instant attenuated virus may be used in a vaccine composition for inducing a protective immune response in a subject. The invention also provides a method of synthesizing the instant attenuated virus. Further, this invention further provides a method for preventing a subject from becoming afflicted with a virus-associated disease comprising administering to the subject a prophylactically effective dose of a vaccine composition comprising the instant attenuated virus.

(56)

References Cited**OTHER PUBLICATIONS**

- Luytjes, W. et al., "Amplification, Expression, and Packaging of a Foreign Gene by Influenza Virus", *Cell* (1989), vol. 59, pp. 1107-1113.
- McKnight, K., "The Human Rhinovirus Internal Cis-acting replication element (cre) Exhibits Disparate Properties among Stereotypes", *Arch Virol*. (2003), vol. 148, pp. 2397-2418.
- Palese, P. et al., Orthomyxoviridae: The Viruses and Their Replication, Ch. 47, pp. 1647-1689 in *Fields Virology* (2007), vol. 2, 5th Edition, David M. Knipe, PhD, Editor-In-Chief, Wolters Kluwer, publisher, Philadelphia, USA.
- Park, S. et al., "Advances in Computational Protein Design", *COSB* (2004), vol. 14, pp. 487-494.
- Paul, A. et al., "Internal Ribosomal Entry Site Scanning of the Poliovirus Polyprotein: Implications for Proteolytic Processing", *Virology* (1998), vol. 250, 241-253.
- Pelletier, J. et al., "Internal Initiation of Translation of Eukaryotic mRNA Directed by a Sequence Derived from Poliovirus RNA", *Nature* (1988), vol. 334, pp. 320-325.
- Rueckert, R.R., "Picornaviruses and Their Replication", Ch. 32, pp. 705-738, in *Virology* (1985), Bernard N. Fields, M.D., Editor-In-Chief, Raven Press, publisher, New York, USA.
- Russell, C. et al., "The Genesis of a Pandemic Influenza Virus", *Cell* (2005), pp. 368-371.
- Savolainen, C. et al., "Human Rhinoviruses", *PRR* (2003), vol. 4, pp. 91-98.
- Tian, J. et al., "Accurate Multiplex Gene Synthesis from Programmable DNA Microchips", *Nature* (2004), vol. 432, pp. 1050-1054.
- Ansardi, D., et al., "Complementation of a Poliovirus Defective Genome by a Recombinant Vaccinia Virus Which Provides Poliovirus P1 Capsid Precursor in Trans", *J. Virol.* (2003), vol. 67:6, pp. 3684-3690.
- Belov, G. et al., "The Major Apoptotic Pathway Activated and Suppressed by Poliovirus", *J. Virol.* (2003), vol. 77:1, pp. 45-56.
- Buchan, J. et al., "tRNA Properties Help Shape Codon Pair Preferences in Open Reading Frames", *Nucl. Acids Res.* (2006), vol. 34:3, pp. 1015-1027.
- Cao, X. et al., "Replication of Poliovirus RNA Containing Two Vpg Coding Sequences Leads to a Specific Deletion Event", *J. Virol.* (1993), vol. 67:9, pp. 5572-5578.
- Carlini, D. et al., "In Vivo Introduction of Unpreferred Synonymous Codons Into the *Drosophila* Adh Gene Results in Reduced Levels of ADH Protein", *Genetics* (2003), vol. 163, pp. 239-243.
- Cello, J. et al., "Chemical Synthesis of Poliovirus CdnA: Generation of Infectious Virus in the Absence of Natural Template", *Science* (2002), vol. 297, pp. 1016-1018.
- Coleman, J.R. et al., "Synthetic Construct Capsid Protein P1-Min Gene, Partial Cds", (2007), retrieved from EBI accession No. EM_{SY}: EU095953; Database accession No. EU095953.
- Coleman, J.R. et al., "Virus Attenuation by Genome-Scale Changes in Codon Pair Bias", *Science* (2008), vol. 320, pp. 1784-1787.
- Corpet, F., "Multiple Sequence Alignment with Hierarchical Clustering", *Nucl. Acids Res.* (1988), vol. 16:22, pp. 10881-10890.
- Crotty, S., et al., "RNA Virus Error Catastrophe: Direct Molecular Test by Using Ribavirin", *Proc. Natl. Acad. Sci. U.S.A.* (2001), vol. 98:12, pp. 6895-6900.
- Curran, J., et al., "Selection of aminoacyl-tRNAs at sense codons: the size of the tRNA variable loop determines whether the immediate 3' nucleotide to the codon has a context effect", *Nucl. Acids Res.* (1995), vol. 23:20, pp. 4104-4108.
- Dove, A., et al., "Cold-Adapted Poliovirus Mutants Bypass a Postentry Replication Block", *J. Virol.* (1997), vol. 71:6, pp. 4728-4735.
- Enami, M. et al., "Introduction of Site-Specific Mutations into the Genome of Influenza Virus", *Proc. Natl. Acad. Sci. U.S.A.* (1990), vol. 87, pp. 3802-3805.
- Farabaugh, P.J. Programmed Translational Frameshifting, *Microbiol Rev.* (1996), vol. 60:1, pp. 103-134.
- Fedorov, A. et al., "Regularities of Context-Dependent Codon Bias in Eukaryotic Genes", *Nucl. Acids Res.* (2002), vol. 30:5, pp. 1192-1197.
- Fodor, E. et al., "Rescue of Influenza A Virus From Recombinant DNA", *J Virol.* (1999), vol. 73:11, pp. 9679-9682.
- Georgescu, M. et al., "Evolution of the Sabin Type 1 Poliovirus in Humans: Characterization of Strains Isolated From Patients with Vaccine-Associated Paralytic Poliomyelitis", *J. Virol.* (1997), vol. 71:10, pp. 7758-7768.
- Gerber, K. et al., "Biochemical and Genetic Studies of the Initiation of Human Rhinovirus 2 RNA Replication: Identification of a Cis-Replicating Element in the Coding Sequence of 2Apro", *J. Virol.* (2001), vol. 75:22, pp. 10979-10990.
- Girard, S. et al., "Poliovirus Induces Apoptosis in the Mouse Central Nervous System", *J. Virol.* (1999), vol. 73:7, pp. 6066-6072.
- Goodfellow, I. et al., "Identification of a Cis-Acting Replication Element Within the Poliovirus Coding Region", *J. Virol.* (2000), vol. 74:10, pp. 4590-4600.
- Gu, W., et al., "Analysis of Synonymous Codon Usage in SARS Coronavirus and other viruses in the Nidovirales", *Virus Research* (2004), vol. 101, pp. 155-161.
- Gutman, G.A., et al., "Nonrandom Utilization of Codon Pairs in *Escherichia coli*", *Proc. Natl. Acad. Sci. U. S. A.* (1989), vol. 86, pp. 3699-3703.
- He, Y. et al., "Interaction of the Poliovirus Receptor with Poliovirus", *Proc. Natl. Acad. Sci. USA* (2000), vol. 97:1, pp. 79-84.
- Herold, J. et al., "Poliovirus Requires a Precise 5' End for Efficient Positive-Strand RNA Synthesis", *J. Virol.* (2001), vol. 74:14, vol. pp. 6394-6400.
- Hoekema, A., et al., "Codon Replacement in the PGK1 Gene of *Saccharomyces cerevisiae*: Experimental Approach to Study the Role of Biased Codon Usage in Gene Expression", *Mol. Cell. Biol.* (1987), vol. 7:8, pp. 2914-2924.
- Hofer, F. et al., "Members of the Low Density Lipoprotein Receptor Family Mediate Cell Entry of a Minor-Group Common Cold Virus", *Proc. Natl. Acad. Sci. U.S.A.* (1994), vol. 91, pp. 1839-1842.
- Hoffmann, E. et al., "A DNA transfection system for generation of influenza: A virus from eight plasmids", *Proc. Natl. Acad. Sci. U.S.A.* (2000), vol. 97:11, pp. 6108-6113.
- Hogle, J. M. "Poliovirus Cell Entry: Common Structural Themes in Viral Cell Entry Pathways", *Annu. Rev. Microbiol.* (2002), vol. 56, pp. 677-702.
- Holland, J.J. et al. "Mutation Frequencies at Defined Single Codon Sites in Vesicular Stomatitis Virus and Poliovirus Can Be Increased Only Slightly by Chemical Mutagenesis", *J. Virol.* (1990), vol. 64:8, pp. 3960-3962.
- Hsiao, L. L., "A Compendium of Gene Expression in Normal Human Tissues", *Physiol. Genomics* (2001), vol. 7, pp. 97-104.
- Irwin, B., et al., "Codon Pair Utilization Biases Influence Translational Elongation Step Times" *J. Biol. Chem.* (1995) vol. 270, pp. 22801-22806.
- Jang, S., et al., "Initiation of Protein Synthesis by Internal Entry of Ribosomes into the 5' Nontranslated Region of Encephalomyocarditis Virus RNA In Vivo", *J. Virol.* (1989), vol. 63, pp. 1651-1660.
- Jayaram, S., et al., "GeMS: an advanced software package for designing synthetic genes", *Nucl. Acids Res.* (2005), vol. 33, pp. 3011-3016.
- Kamps, B., et al., *Influenza Report*, (2006) (eds.), Flying Publisher, 225 pgs.
- Kaplan, G., et al., "Construction and Characterization of Poliovirus Subgenomic Replicons", *J. Virol.* (1988), vol. 62, pp. 1687-1696.
- Karlin, S., et al., "Why is Cpg Suppressed in the Genomes of Virtually All Small Eukaryotic Viruses but not in Those of Large Eukaryotic Viruses?", *J. Virol.* (1994), vol. 68, pp. 2889-2897.
- Kew, O., et al., "Outbreak of Poliomyelitis in Hispaniola Associated With Circulating Type 1 Vaccine-Derived Poliovirus", *Science* (2002) vol. 296, pp. 356-359.
- Kilbourne, E.D., "Influenza pandemics of the 20th century", *Emerg. Infect. Dis.* (2006), vol. 12, pp. 9-14.
- Koike, S., et al., "Transgenic Mice Susceptible to Poliovirus", *Proc. Natl. Acad. Sci.* (1991), vol. 88, pp. 951-955.

(56)

References Cited**OTHER PUBLICATIONS**

- Ledford, R.M., et al., "VP1 Sequencing of All Human Rhinovirus Serotypes: Insights Into Genus Phylogeny and Susceptibility to Antiviral Capsid-Binding Compounds", *J. Virol.* (2004), vol. 78, pp. 3663-3674.
- Molla, A., et al., "Cell-Free, De Novo Synthesis of Poliovirus", *Science* (1991), vol. 254, pp. 1647-1651.
- Mueller, S., et al., "Poliovirus and poliomyelitis: a tale of guts, brains, and an accidental event" *Virus Res.* (2005), vol. 111, pp. 175-193.
- Murdin, A., et al., "Construction of a poliovirus type 1/type 2 antigenic hybrid by manipulation of neutralization antigenic site II", *J. Virol.* (1989), vol. 63, pp. 5251-5257.
- Neumann, G., et al., "Generation of Influenza A Viruses Entirely From Clone Cdnas", *Proc. Natl. Acad. Sci.*, vol. 96, pp. 9345-9350, 1999.
- Neznanov, N., et al., "Proteolytic Cleavage of the P65-RelA Subunit of NF-Kappab During Poliovirus Infection", *J. Biol. Chem.* (2005), vol. 280, pp. 24153-24158.
- Pfister, T., et al., "Characterization of the Nucleoside Triphosphatase Activity of Poliovirus Protein 2C Reveals a Mechanism by Which Guanidine Inhibits Poliovirus Replication", *J. Biol. Chem.* (1999), vol. 274, pp. 6992-7001.
- Plotkin, J., et al., "Tissue-Specific Codon Usage and the Expression of Human Genes", *Proc. Natl. Acad. Sci.* (2004), vol. 101, pp. 12588-12591.
- Racaniello, V., et al., "Cloned Poliovirus Complementary DNA is Infectious in Mammalian Cells", *Science* (1981), vol. 214, pp. 916-919.
- Richardson, S., et al., "Genedesign: Rapid, Automated Design of Multikilobase Synthetic Genes", *Genome Res.* (2006), vol. 16, pp. 550-556.
- Robinson, M., "Codon Usage Can Affect Efficiency of Translation of Genes in *Escherichia coli*.", *Nucl. Acids Res.* (2006), vol. 12, pp. 6663-6671.
- Sánchez, G., et al., "Genome Variability and Capsid Structural Constraints of Hepatitis A Virus", *J. Virol.* (2003), vol. 77, pp. 452-459.
- Shimizu, H., et al., "Circulation of Type 1 Vaccine-Derived Poliovirus in the Philippines in 2001", *J. Virol.* (2004), vol. 78, pp. 13512-13521.
- Simonsen, L., "Impact of Influenza Vaccination on Seasonal Mortality in the US Elderly Population", *Arch. Intern. Med.* (2005), vol. 165, pp. 265-272.
- Skiena, S. "Designing Better Phages" *Bioinformatics* (2001), vol. 17 Suppl 1, pp. 5253-5261.
- Stephenson, I., et al., "Influenza: Current Threat from Avian Influenza", *Br. Med. Bull.* (2005), vol. 75-76, pp. 63-80.
- Talon, J., et al., "Influenza A and B Viruses Expressing Altered NS1 Proteins: A Vaccine Approach", *Proc. Natl. Acad. Sci.* (2000), vol. 97, pp. 4309-4314.
- Thompson, W., et al., "Mortality associated with influenza and respiratory syncytial virus in the United States", *JAMA*. (2003), vol. 289, pp. 179-186.
- Tolskaya, E., et al., "Apoptosis-Inducing and Apoptosis-Preventing Functions of Poliovirus", *J. Virol.* (1995), vol. 69, pp. 1181-1189.
- Toyoda, H., et al., "Oncolytic Treatment and Cure of Neuroblastoma by a Novel Attenuated Poliovirus in a Novel Poliovirus-Susceptible Animal Model" *Cancer Res.* (2007), vol. 67, pp. 2857-2864.
- Van Der Wert, S., et al., "Synthesis of infectious poliovirus RNA by purified T7 RNA polymerase", *Proc. Natl. Acad. Sci.* (1986), vol. 78, pp. 2330-2334.
- Wahby, A., "Combined Cell Culture Enzyme-Linked Immunosorbent Assay for Quantification of Poliovirus Neutralization-Relevant Antibodies", *Clin. Diagn. Lab. Immunol.* (2000) vol. 7, pp. 915-919.
- Wang, B., et al., Two Proteins for the Price of One: The Design of Maximally Compressed Coding Sequences Natural Computing. Eleventh International Meeting on DNA Based Computers (DNA11), 2005. Lecture Notes in Computer Science (LNCS), (2006) vol. 3892, pp. 387-398.
- Wimmer, E., et al., "Synthetic Viruses: A New Opportunity to Understand and Prevent Viral Disease", *Nat. Biotech.* (2009), vol. 27:12, pp. 1163-1172.
- Zhao, W.D., et al., "Genetic Analysis of a Poliovirus/Hepatitis C Virus Chimera: New Structure for Domain II of the Internal Ribosomal Entry Site of Hepatitis C Virus", *J. Virol.* (2001) vol. 75, pp. 3719-3730.
- Zhou, J., et al., "Papillomavirus Capsid Protein Expression Level Depends on the Match Between Codon Usage and tRNA Availability", *J. Virol.* (1999), vol. 73, pp. 4972-4982.
- Zhou, T. et al., "Analysis of Synonymous Codon Usage in H5N1 Virus and Other Influenza A Viruses", *Biosystems* (2005), vol. 81:1, pp. 77-86.
- Zolotukhin, S., et al., "A "Humanized" Green Fluorescent Protein Cdnas Adapted for High-Level Expression in Mammalian Cells", *J. Virol.*, (1996), vol. 70, pp. 4646-4654.

* cited by examiner

AA	Codon	Human Freq.	PV(M)	PV-SD	PV-AB
Ala	GCT	26.3%	14	14	3
	GCC	40.3%	17	17	1
	GCA	22.6%	22	22	2
	GCG	10.8%	11	11	58
Cys	TGT	45.0%	8	8	15
	TGC	55.0%	7	7	0
Asp	GAT	46.2%	22	23	46
	GAC	53.8%	26	25	2
Glu	GAA	41.9%	18	17	32
	GAG	58.1%	16	17	2
Phe	TTT	45.5%	14	14	33
	TTC	54.5%	21	21	2
Gly	GGT	16.3%	15	16	51
	GGC	34.1%	10	9	1
	GGA	24.7%	15	15	0
	GGG	24.9%	13	13	1
His	CAT	41.2%	9	9	18
	CAC	58.8%	10	10	1
Ile	ATT	35.8%	13	13	0
	ATC	47.9%	13	13	0
	ATA	16.3%	15	15	41
Lys	AAA	42.6%	17	17	34
	ARG	57.4%	18	18	1
Leu	TTR	7.4%	10	9	66
	TIG	12.7%	13	14	2
	CTT	13.0%	10	10	0
	CTC	19.7%	9	9	0
	CTA	7.0%	11	11	2
	CTG	40.2%	17	17	0
Met	ATG	100.0%	25	25	25
Asn	AAI	46.5%	25	24	49
	AAC	53.5%	25	26	1
Pro	CCT	28.3%	17	16	1
	CCC	32.7%	8	9	1
	CCA	27.4%	27	27	1
	CCG	11.6%	10	10	59
Gln	CAA	25.5%	13	13	28
	CAG	74.1%	16	16	1
Arg	CGT	8.1%	4	4	36
	CGC	18.9%	3	3	0
	CGA	11.0%	2	2	1
	CGG	20.7%	6	6	0
	AGA	20.7%	12	12	0
	AGG	20.6%	11	11	1
Ser	TCT	18.5%	10	11	2
	TCC	22.0%	16	15	0
	TCA	14.8%	19	19	0
	TCG	5.6%	8	8	69
	AGT	14.9%	8	8	0
	AGC	24.2%	10	10	0
Thr	ACT	24.4%	16	16	0
	ACC	36.0%	35	21	1
	ACA	28.0%	21	21	0
	ACG	11.6%	10	10	81
Val	GTT	17.8%	6	6	0
	GTC	24.0%	13	13	3
	GTA	11.4%	14	14	0
	GTG	46.8%	25	25	55
Trp	TGG	100.0%	13	13	13
	TAT	43.8%	18	19	37
Tyr	TAC	56.2%	21	20	2

Fig. 1

Fig. 2a

Fig. 2b

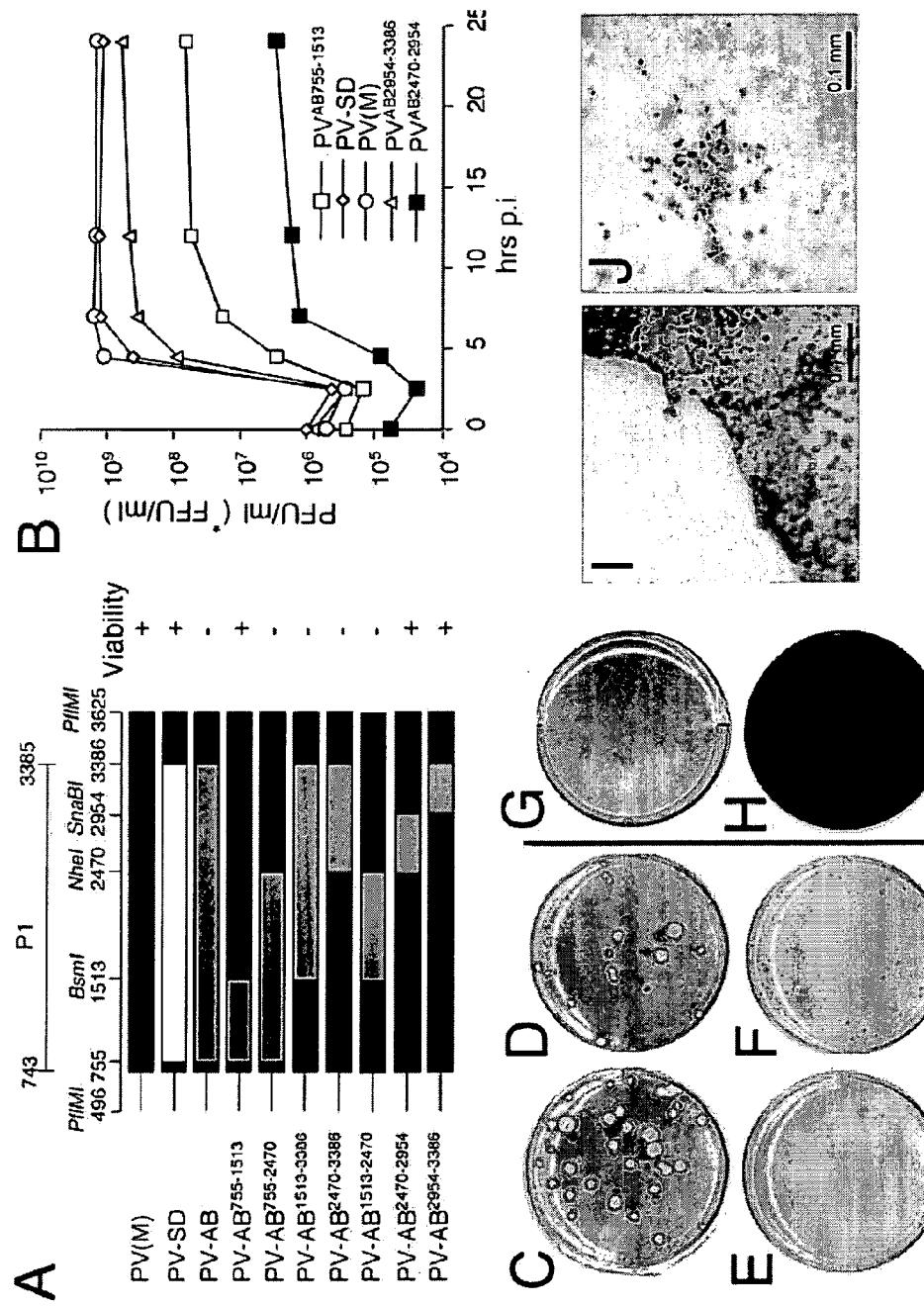


Fig. 3

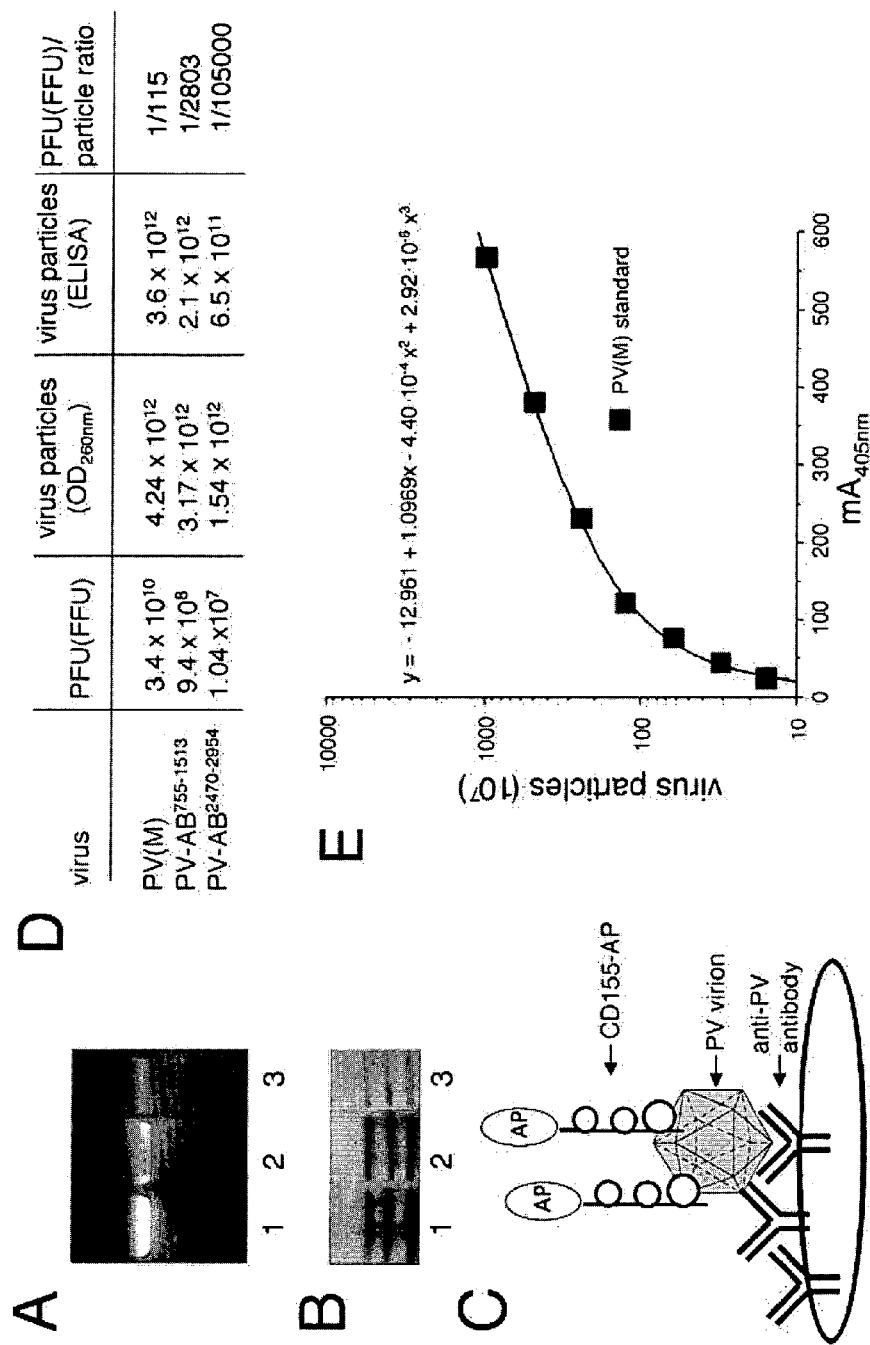


Fig. 4

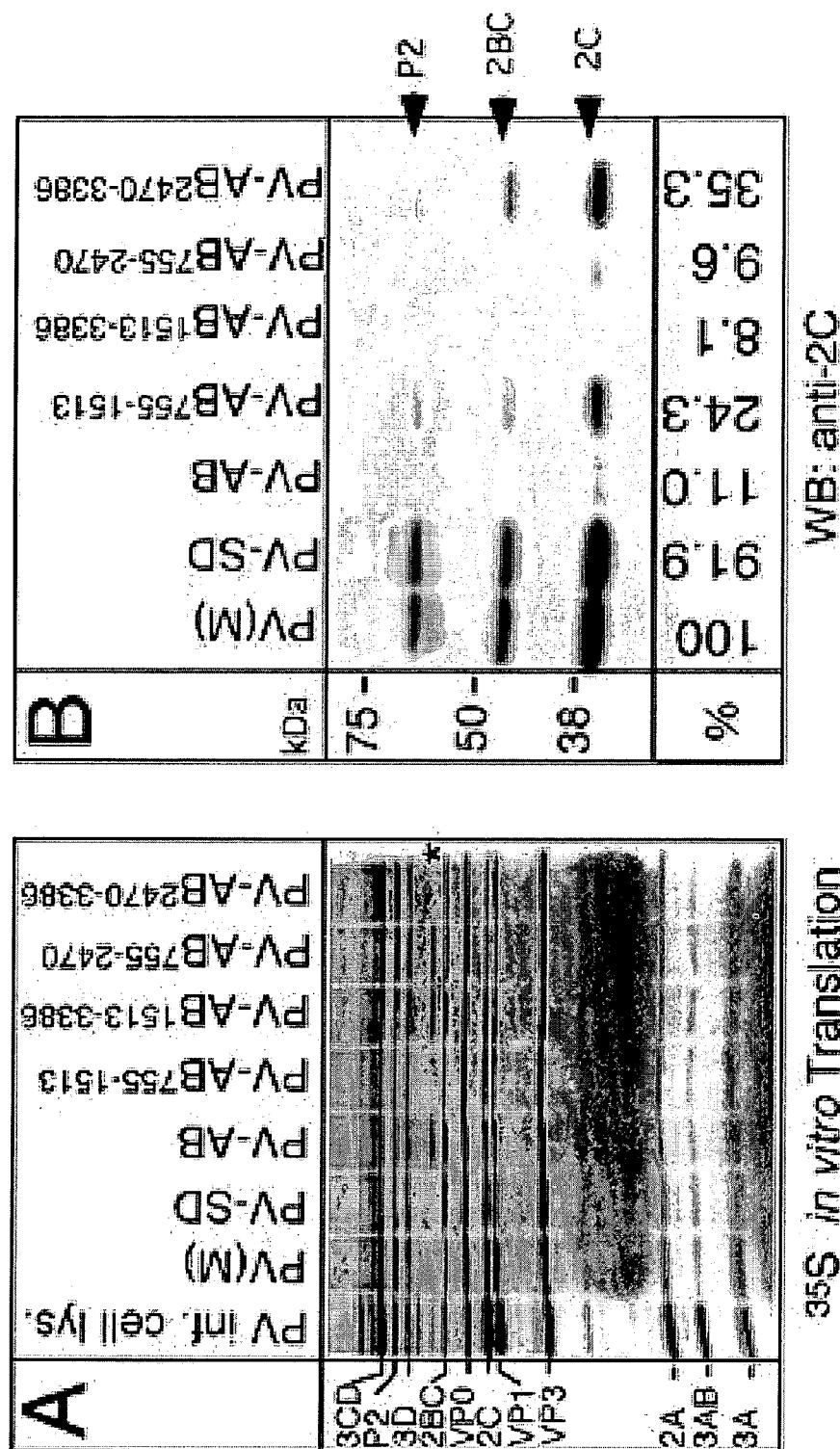


Fig. 5

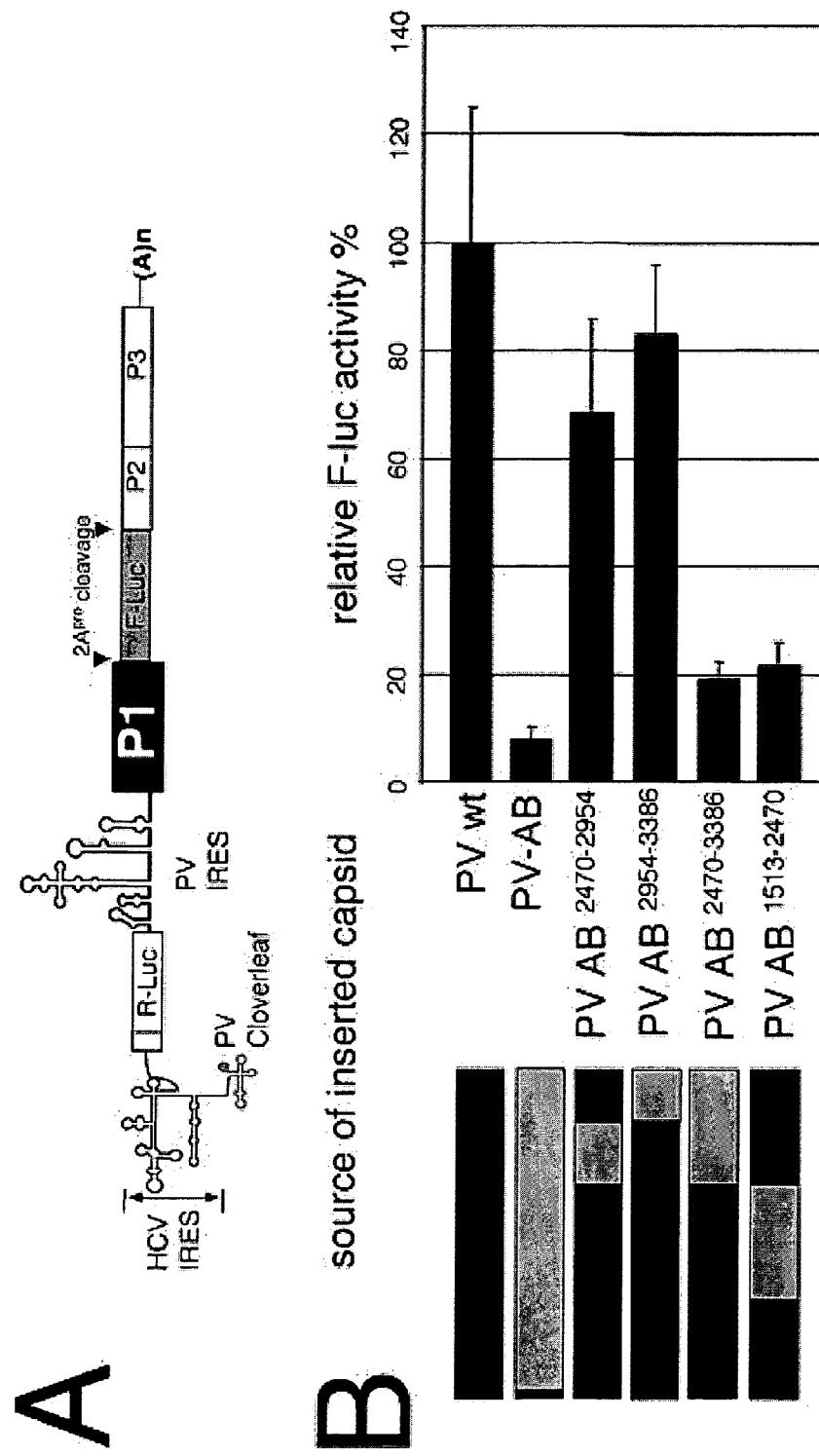


Fig. 6

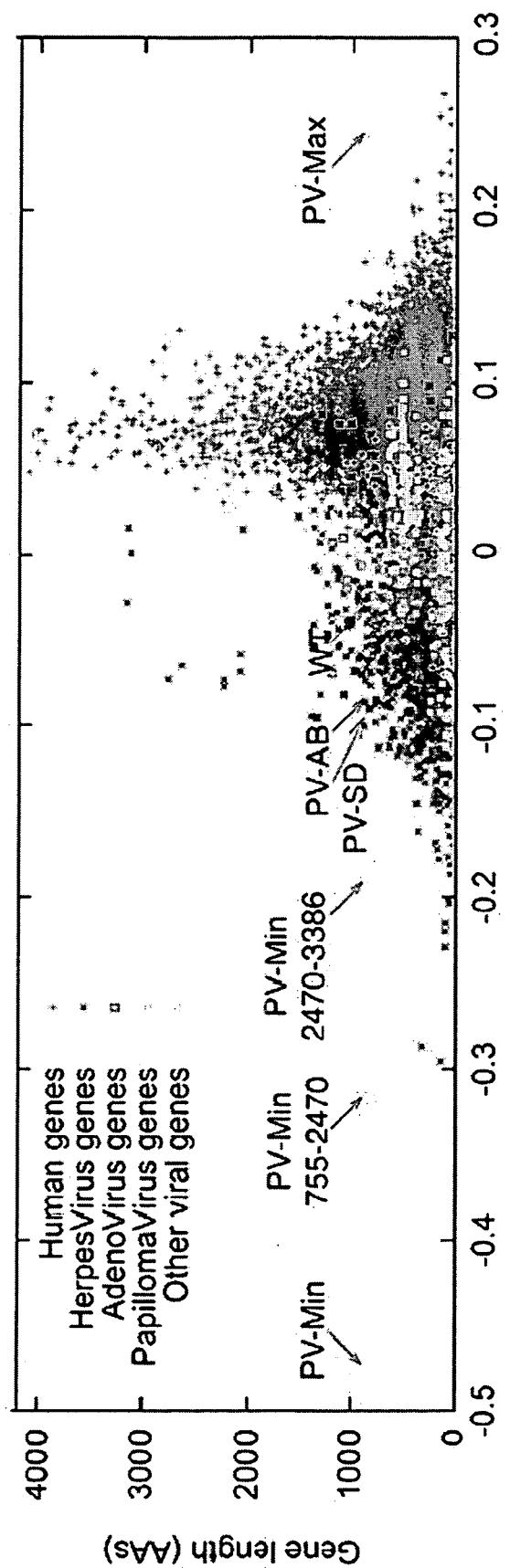


Fig. 7

Virus	Titer PFU/ml	PFU/Particle Ratio
PV(M)	1.2×10^9	1/115
PV-AB ^{755-T513}	6.7×10^7	1/2803
PV-AB ²⁴⁷⁰⁻²⁹⁵⁴	3×10^6	1/105000
PV-Min ⁷⁵⁵⁻²⁴⁷⁰	2.4×10^6	1/71000
PV-Min ²⁴⁷⁰⁻³³⁸⁶	1.8×10^6	1/17500
PV-Max	2.3×10^9	N/D

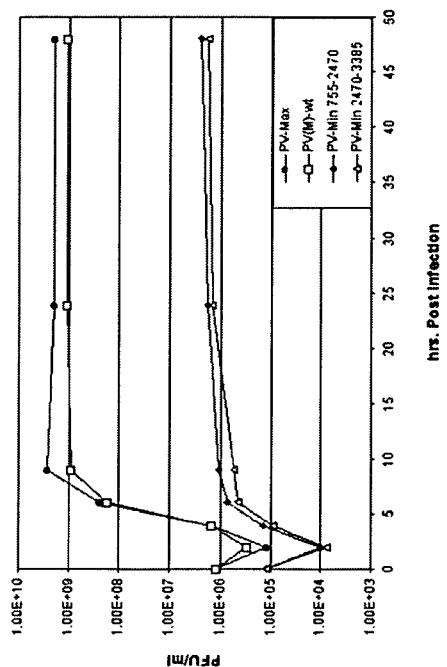


Fig. 8

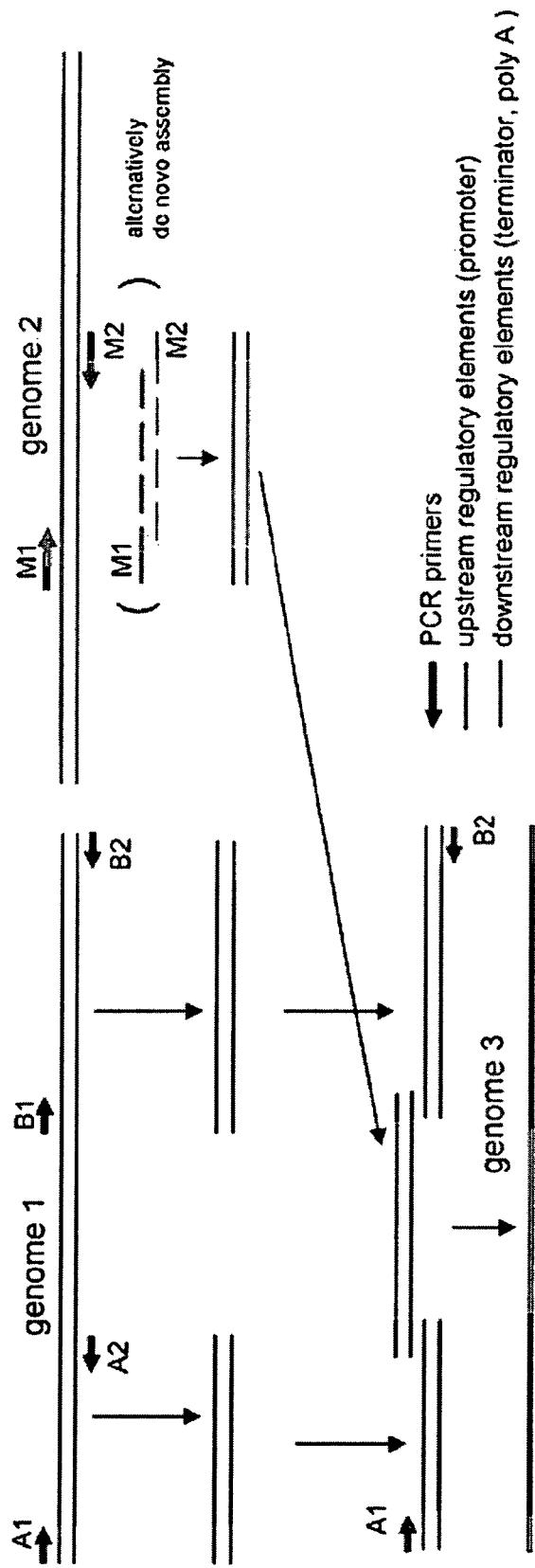


Fig. 9

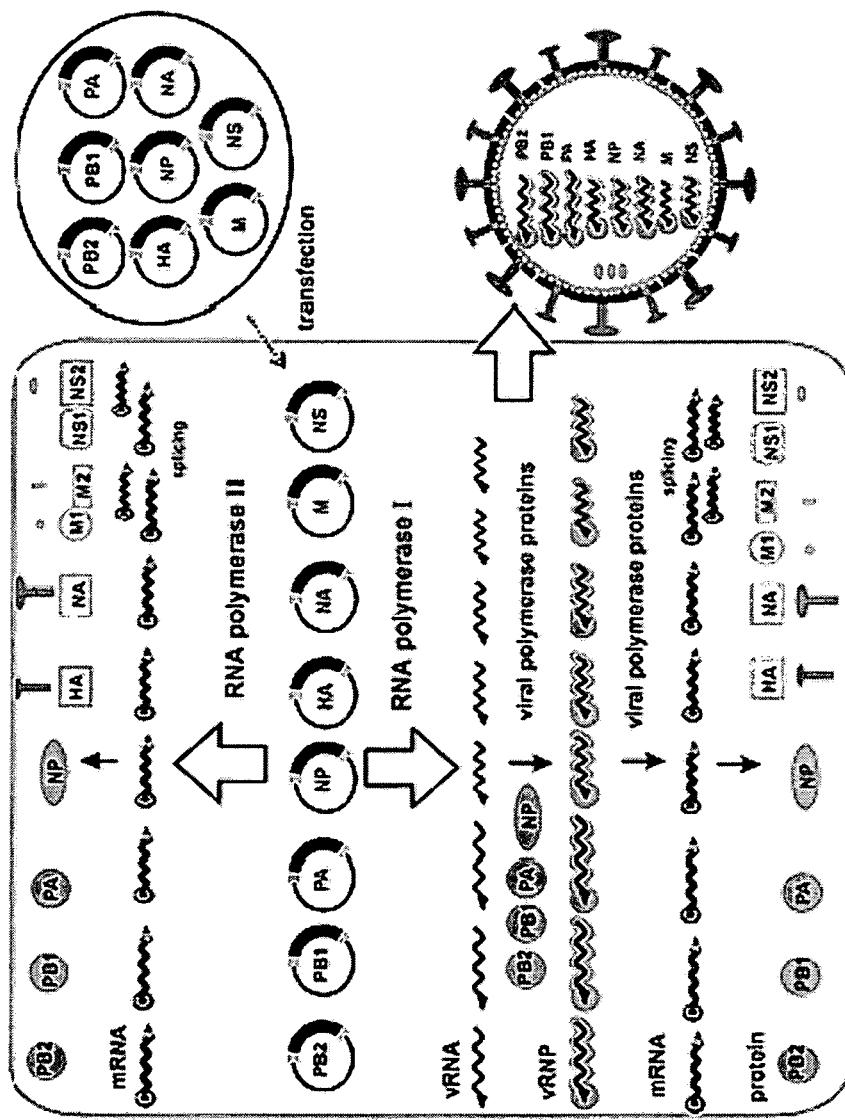


Fig. 10

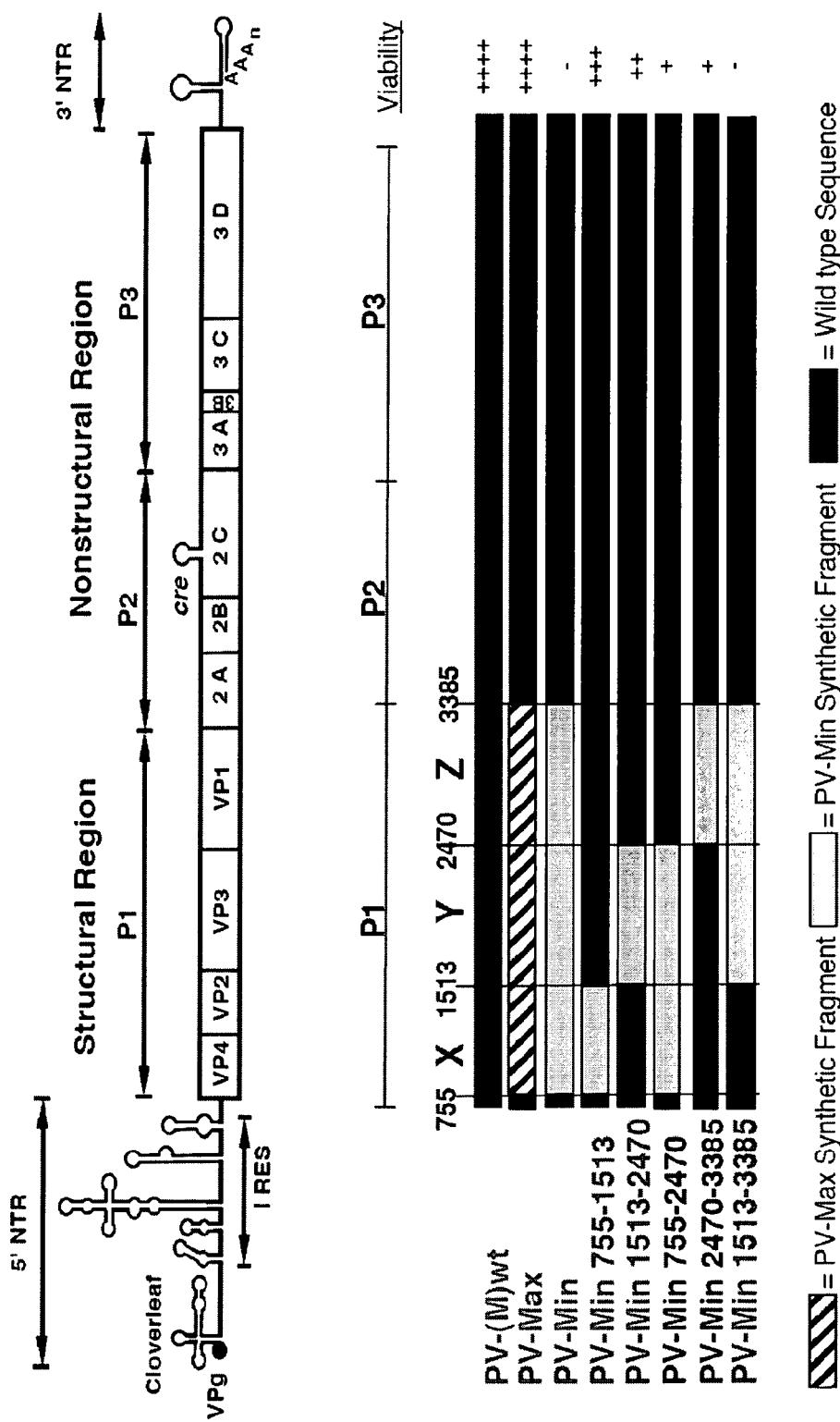


FIG. 11

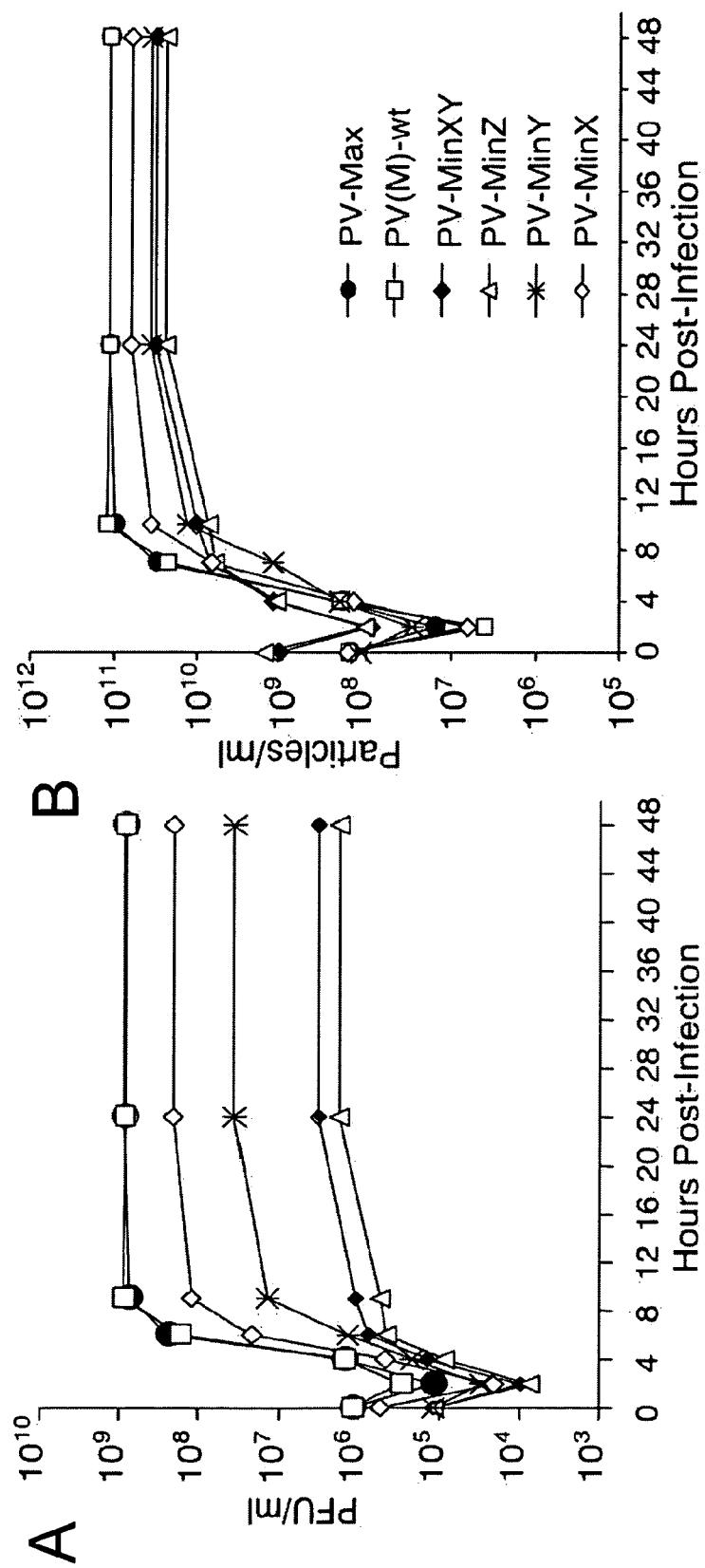


Fig. 12

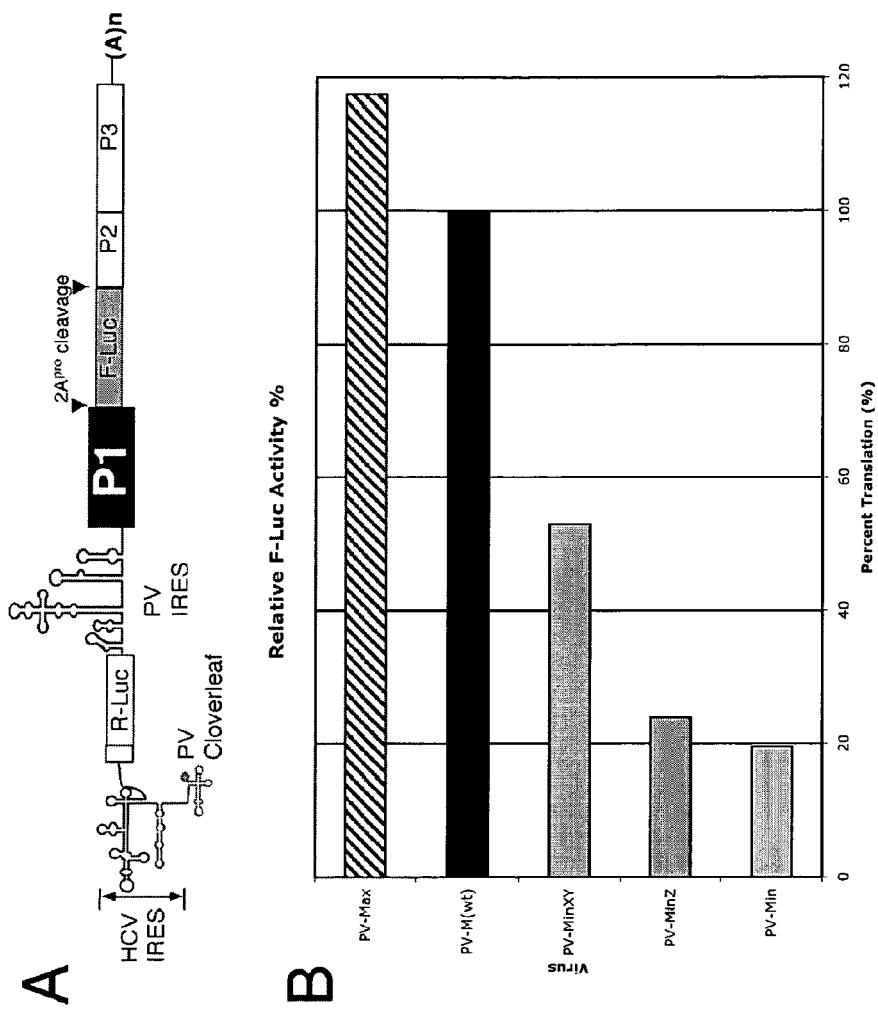


Fig. 13

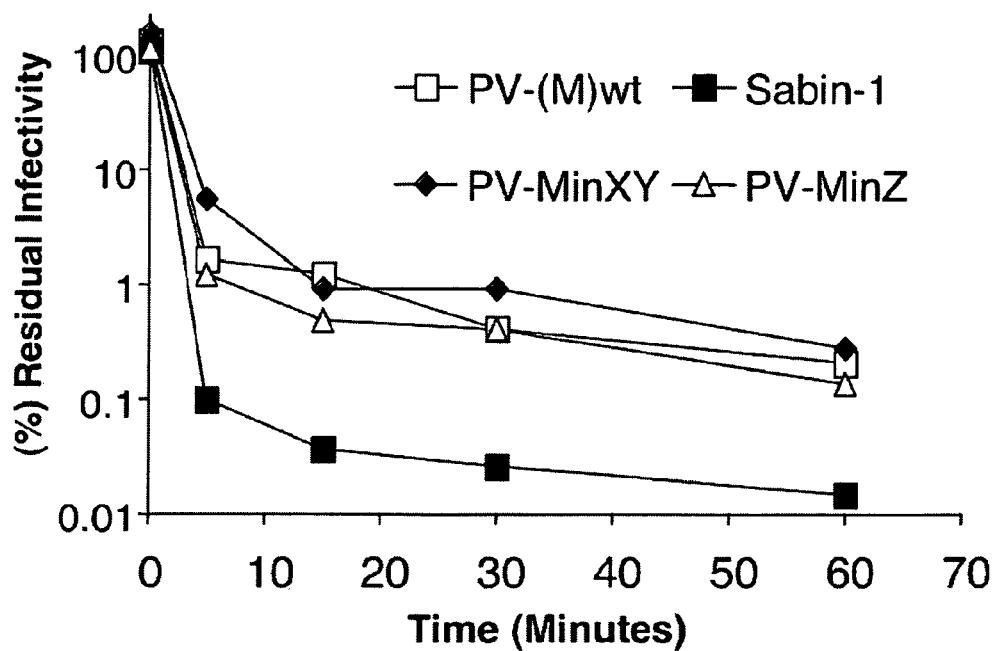


Fig. 14

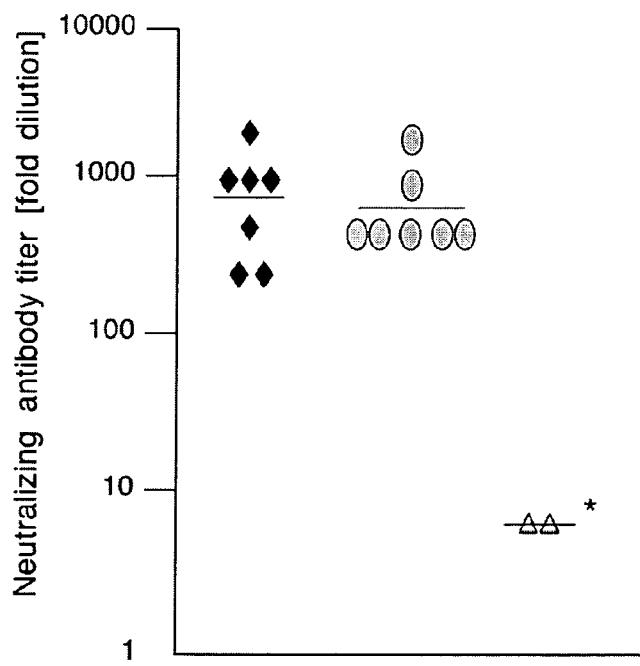


Fig. 15

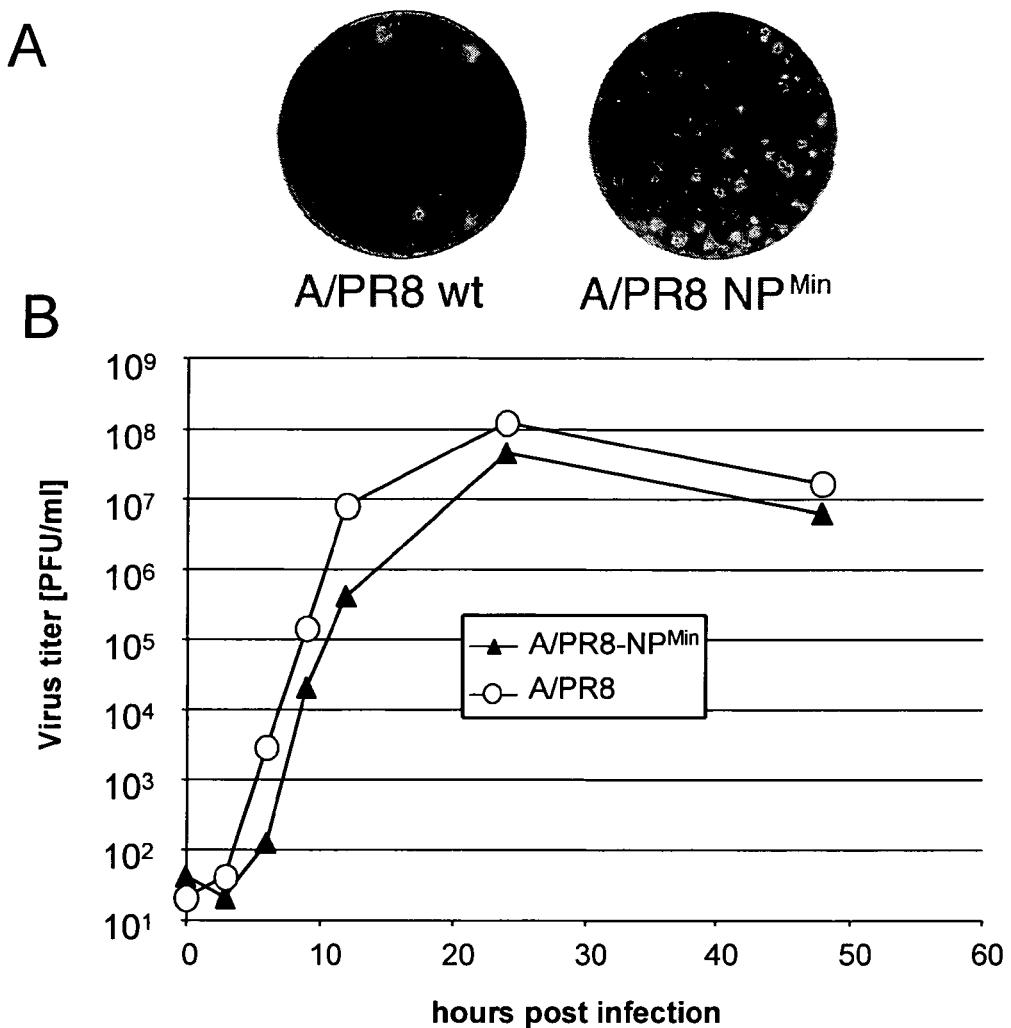


Fig. 16

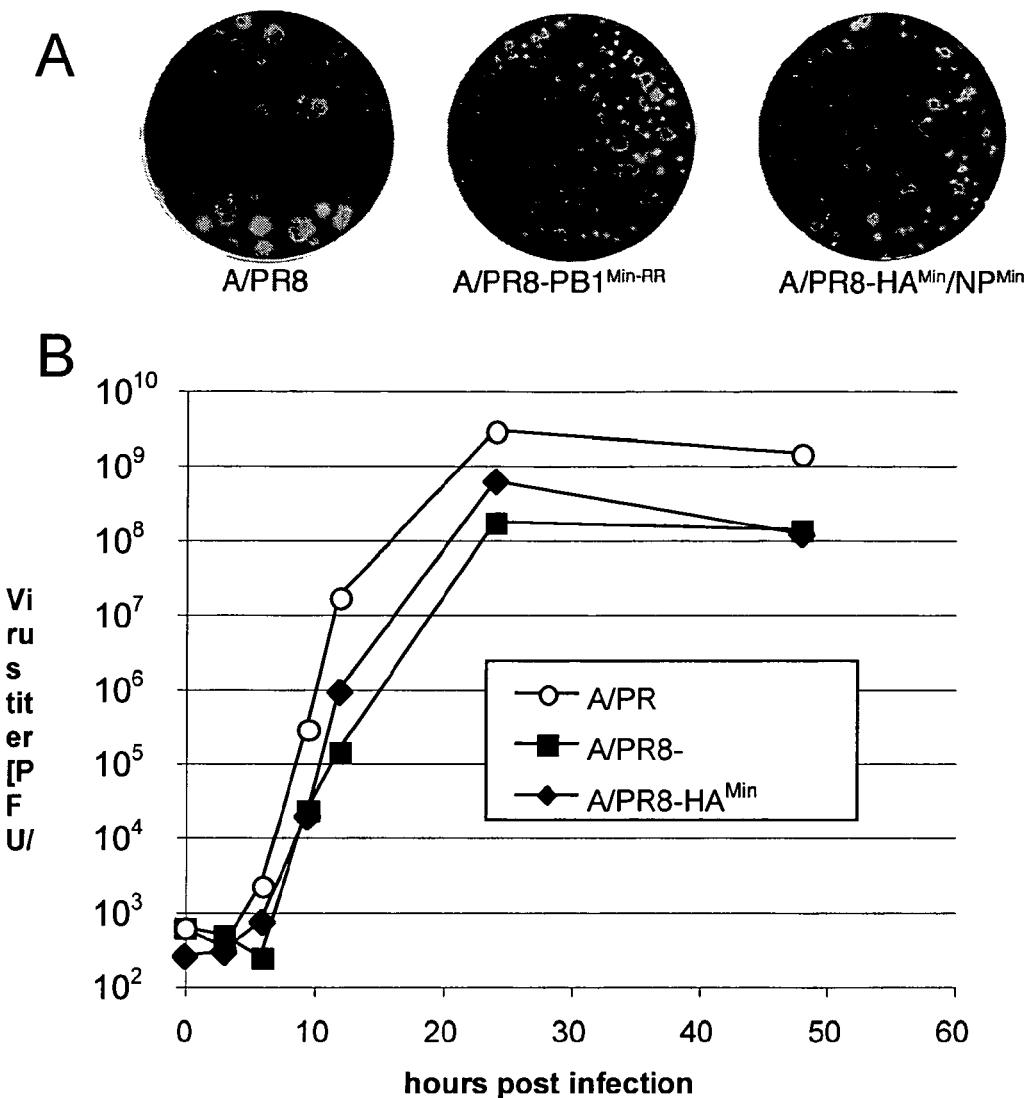


Fig. 17

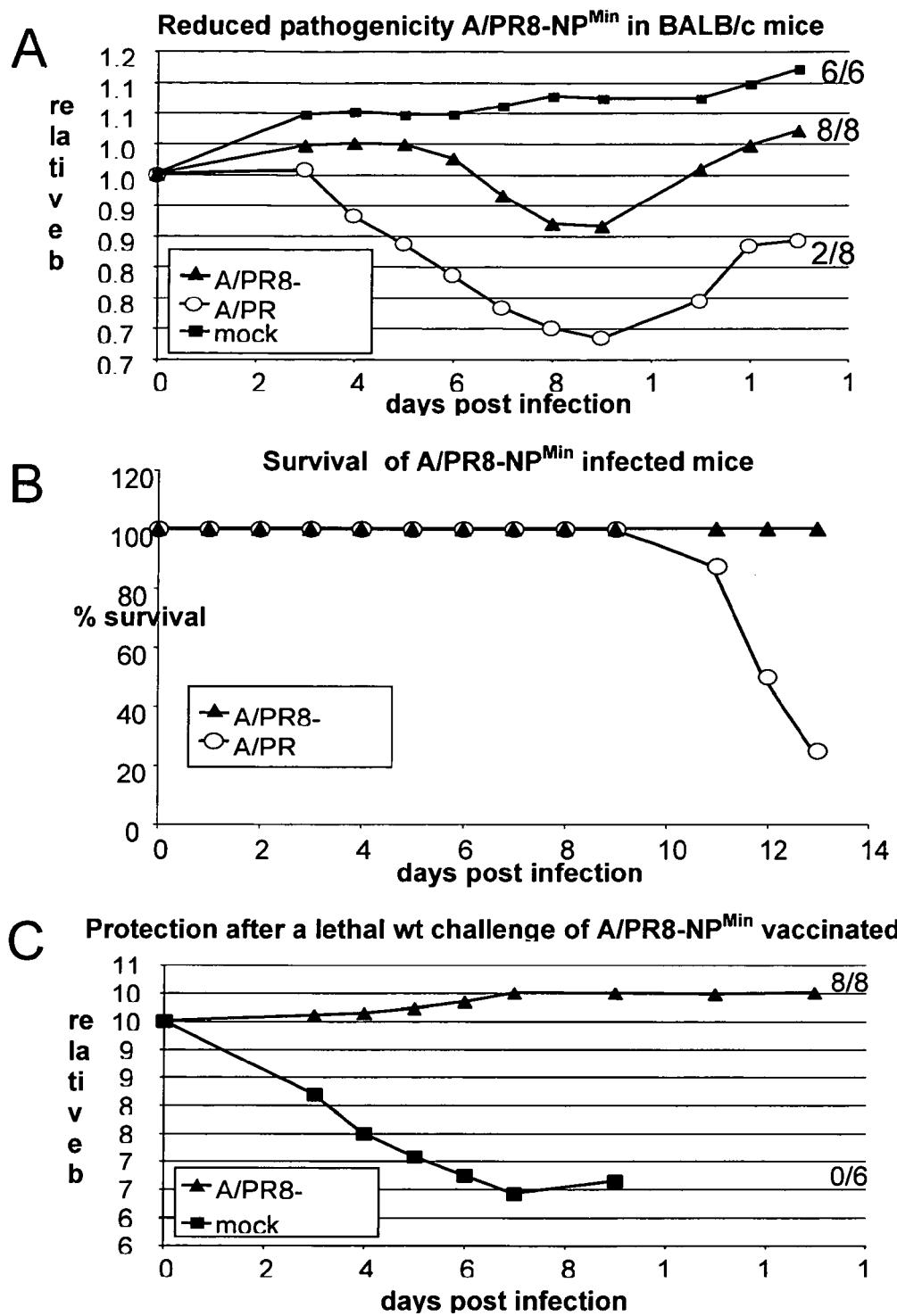


Fig. 18

1**ATTENUATED VIRUSES USEFUL FOR VACCINES****CROSS REFERENCE TO RELATED APPLICATIONS**

This application is the national phase application of International application number PCT/US2008/058952 filed Mar. 31, 2008, which claims the benefit of priority to U.S. Application No. 60/909,389, filed Mar. 30, 2007, and U.S. Application No. 61/068,666, filed Mar. 7, 2008, which are incorporated herein by reference in their entireties.

FEDERAL FUNDING

The invention was made with United States Government support under Grant Nos. AI15122 from the National Institutes of Health, EIA0325123 from the National Science Foundation, and T32-CA009176 from the National Cancer Institute. Accordingly, the United States Government has certain rights in this invention.

COPYRIGHT NOTICE

A portion of the disclosure of this patent document contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or patent disclosure as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

FIELD OF THE INVENTION

The present invention relates to the creation of an attenuated virus comprising a modified viral genome containing a plurality of nucleotide substitutions. The nucleotide substitutions result in the exchange of codons for other synonymous codons and/or codon rearrangement and variation of codon pair bias.

BACKGROUND OF THE INVENTION

Rapid improvements in DNA synthesis technology promise to revolutionize traditional methods employed in virology. One of the approaches traditionally used to eliminate the functions of different regions of the viral genome makes extensive but laborious use of site-directed mutagenesis to explore the impact of small sequence variations in the genomes of virus strains. However, viral genomes, especially of RNA viruses, are relatively short, often less than 10,000 bases long, making them amenable to whole genome synthesis using currently available technology. Recently developed microfluidic chip-based technologies can perform *de novo* synthesis of new genomes designed to specification for only a few hundred dollars each. This permits the generation of entirely novel coding sequences or the modulation of existing sequences to a degree practically impossible with traditional cloning methods.

Such freedom of design provides tremendous power to perform large-scale redesign of DNA/RNA coding sequences to: (1) study the impact of changes in parameters such as codon bias, codon-pair bias, and RNA secondary structure on viral translation and replication efficiency; (2) perform efficient full genome scans for unknown regulatory elements and other signals necessary for successful viral

2

reproduction; and (3) develop new biotechnologies for genetic engineering of viral strains and design of anti-viral vaccines.

As a result of the degeneracy of the genetic code, all but two amino acids in the protein coding sequence can be encoded by more than one codon. The frequencies with which such synonymous codons are used are unequal and have coevolved with the cell's translation machinery to avoid excessive use of suboptimal codons that often correspond to rare or otherwise disadvantaged tRNAs (Gustafsson et al., 2004). This results in a phenomenon termed "synonymous codon bias," which varies greatly between evolutionarily distant species and possibly even between different tissues in the same species (Plotkin et al., 2004).

Codon optimization by recombinant methods (that is, to bring a gene's synonymous codon use into correspondence with the host cell's codon bias) has been widely used to improve cross-species expression (see, e.g., Gustafsson et al., 2004). Though the opposite objective of reducing expression by intentional introduction of suboptimal synonymous codons has not been extensively investigated, isolated reports indicate that replacement of natural codons by rare codons can reduce the level of gene expression in different organisms. See, e.g., Robinson et al., 1984; Hoekema et al., 1987; Carlini and Stephan, 2003; Zhou et al., 1999. Accordingly, the introduction of deoptimized synonymous codons into a viral genome may adversely affect protein translation and thereby provide a method for producing attenuated viruses that would be useful for making vaccines against viral diseases.

Viral Disease and Vaccines

Viruses have always been one of the main causes of death and disease in man. Unlike bacterial diseases, viral diseases are not susceptible to antibiotics and are thus difficult to treat. Accordingly, vaccination has been humankind's main and most robust defense against viruses. Today, some of the oldest and most serious viral diseases such as smallpox and poliomyelitis (polio) have been eradicated (or nearly so) by world-wide programs of immunization. However, many other old viruses such as rhinovirus and influenza virus are poorly controlled, and still create substantial problems, though these problems vary from year to year and country to country. In addition, new viruses, such as Human Immunodeficiency Virus (HIV) and Severe Acute Respiratory Syndrome (SARS) virus, regularly appear in human populations and often cause deadly pandemics. There is also potential for lethal man-made or man-altered viruses for intentional introduction as a means of warfare or terrorism.

Effective manufacture of vaccines remains an unpredictable undertaking. There are three major kinds of vaccines: subunit vaccines, inactivated (killed) vaccines, and attenuated live vaccines. For a subunit vaccine, one or several proteins from the virus (e.g., a capsid protein made using recombinant DNA technology) are used as the vaccine. Subunit vaccines produced in *Escherichia coli* or yeast are very safe and pose no threat of viral disease. Their efficacy, however, can be low because not all of the immunogenic viral proteins are present, and those that are present may not exist in their native conformations.

Inactivated (killed) vaccines are made by growing more-or-less wild type (wt) virus and then inactivating it, for instance, with formaldehyde (as in the Salk polio vaccine). A great deal of experimentation is required to find an inactivation treatment that kills all of the virus and yet does not damage the immunogenicity of the particle. In addition,

residual safety issues remain in that the facility for growing the virus may allow virulent virus to escape or the inactivation may fail.

An attenuated live vaccine comprises a virus that has been subjected to mutations rendering it less virulent and usable for immunization. Live, attenuated viruses have many advantages as vaccines: they are often easy, fast, and cheap to manufacture; they are often easy to administer (the Sabin polio vaccine, for instance, was administered orally on sugar cubes); and sometimes the residual growth of the attenuated virus allows "herd" immunization (immunization of people in close contact with the primary patient). These advantages are particularly important in an emergency, when a vaccine is rapidly needed. The major drawback of an attenuated vaccine is that it has some significant frequency of reversion to wt virulence. For this reason, the Sabin vaccine is no longer used in the United States.

Accordingly, there remains a need for a systematic approach to generating attenuated live viruses that have practically no possibility of reversion and thus provide a fast, efficient, and safe method of manufacturing a vaccine. The present invention fulfills this need by providing a systematic approach, Synthetic Attenuated Virus Engineering (SAVE), for generating attenuated live viruses that have essentially no possibility of reversion because they contain hundreds or thousands of small defects. This method is broadly applicable to a wide range of viruses and provides an effective approach for producing a wide variety of anti-viral vaccines.

SUMMARY OF THE INVENTION

The present invention provides an attenuated virus which comprises a modified viral genome containing nucleotide substitutions engineered in multiple locations in the genome, wherein the substitutions introduce a plurality of synonymous codons into the genome. This substitution of synonymous codons alters various parameters, including codon bias, codon pair bias, density of deoptimized codons and deoptimized codon pairs, RNA secondary structure, CpG dinucleotide content, C+G content, translation frameshift sites, translation pause sites, the presence or absence of tissue specific microRNA recognition sequences, or any combination thereof, in the genome. Because of the large number of defects involved, the attenuated virus of the invention provides a means of producing stably attenuated, live vaccines against a wide variety of viral diseases.

In one embodiment, an attenuated virus is provided which comprises a nucleic acid sequence encoding a viral protein or a portion thereof that is identical to the corresponding sequence of a parent virus, wherein the nucleotide sequence of the attenuated virus contains the codons of a parent sequence from which it is derived, and wherein the nucleotide sequence is less than 90% identical to the nucleotide sequence of the parent virus. In another embodiment, the nucleotide sequence is less than 80% identical to the sequence of the parent virus. The substituted nucleotide sequence which provides for attenuation is at least 100 nucleotides in length, or at least 250 nucleotides in length, or at least 500 nucleotides in length, or at least 1000 nucleotides in length. The codon pair bias of the attenuated sequence is less than the codon pair bias of the parent virus, and is reduced by at least about 0.05, or at least about 0.1, or at least about 0.2.

The virus to be attenuated can be an animal or plant virus. In certain embodiments, the virus is a human virus. In another embodiment, the virus infects multiple species.

Particular embodiments include, but are not limited to, poliovirus, influenza virus, Dengue virus, HIV, rotavirus, and SARS.

This invention also provides a vaccine composition for inducing a protective immune response in a subject comprising the instant attenuated virus and a pharmaceutically acceptable carrier. The invention further provides a modified host cell line specially engineered to be permissive for an attenuated virus that is inviable in a wild type host cell.

In addition, the subject invention provides a method of synthesizing the instant attenuated virus comprising (a) identifying codons in multiple locations within at least one non-regulatory portion of the viral genome, which codons can be replaced by synonymous codons; (b) selecting a synonymous codon to be substituted for each of the identified codons; and (c) substituting a synonymous codon for each of the identified codons.

Moreover, the subject invention provides a method of synthesizing the instant attenuated virus comprising changing the order, within the coding region, of existing codons encoding the same amino acid in order to modulate codon pair bias.

Even further, the subject invention provides a method of synthesizing the instant attenuated virus that combines the previous two methods.

According to the invention, attenuated virus particles are made by transfecting viral genomes into host cells, whereby attenuated virus particles are produced. The invention further provides pharmaceutical compositions comprising attenuated virus which are suitable for immunization.

This invention further provides methods for eliciting a protective immune response in a subject, for preventing a subject from becoming afflicted with a virus-associated disease, and for delaying the onset, or slowing the rate of progression, of a virus-associated disease in a virus-infected subject, comprising administering to the subject a prophylactically or therapeutically effective dose of the instant vaccine composition.

The present invention further provides an attenuated virus which comprises a modified viral genome containing nucleotide substitutions engineered in multiple locations in the genome, wherein the substitutions introduce a plurality of synonymous codons into the genome, wherein the nucleotide substitutions are selected by a process comprising the steps of initially creating a coding sequence by randomly assigning synonymous codons in respective amino acid allowed positions, calculating a codon pair score of the coding sequence randomly selecting and exchanging either (a) pairs of codons encoding the same amino acids or (b) substituting synonymous codons in accordance with a simulated annealing optimization function and repeating the previous step until no further improvement (no change in pair score or bias) is observed for a specific or sufficient number of iterations, until the solution converges on an optima or near optimal value

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1. Codon use statistics in synthetic P1 capsid designs. PV-SD maintains nearly identical codon frequencies compared to wt, while maximizing codon positional changes within the sequence. In PV-AB capsids, the use of nonpreferred codons was maximized. The lengths of the bars and the numbers behind each bar indicate the occurrence of each codon in the sequence. As a reference, the normal

human synonymous codon frequencies ("Freq." expressed as a percentage) for each amino acid are given in the third column.

FIG. 2. Sequence alignment of PV(M), PV-AB and PV-SD capsid coding regions. The nucleotide sequences of PV(M) (SEQ ID NO:1), PV-AB (SEQ ID NO:2) and PV-SD (SEQ ID NO:3) were aligned using the MultAlin online software tool (Corpet, 1988). Numbers above the sequence refer to the position within the capsid sequence. Nucleotide 1 corresponds to nucleotide 743 in the PV(M) virus genome. In the consensus sequence, the occurrence of the same nucleotide in all three sequences is indicated by an upper case letter; the occurrence of the same nucleotide in two of the three sequences is indicated by a lower case letter; and the occurrence of three different nucleotides in the three sequences is indicated by a period.

FIG. 3. Codon-deoptimized virus phenotypes. (A) Overview of virus constructs used in this study. (B) One-step growth kinetics in HeLa cell monolayers. (C to H) Plaque phenotypes of codon-deoptimized viruses after 48 h (C to F) or 72 h (G and H) of incubation; stained with anti-3D^{pol} antibody to visualize infected cells. (C) PV(M), (D) PV-SD, (E) PV-AB, (F) PV-AB⁷⁵⁵⁻¹⁵¹³, (G and H) PV-AB²⁴⁷⁰⁻²⁹⁵⁴. Cleared plaque areas are outlined by a rim of infected cells (C and D). (H) No plaques are apparent with PV-AB²⁴⁷⁰⁻²⁹⁵⁴ after subsequent crystal violet staining of the well shown in panel G. (I and J) Microphotographs of the edge of an immunostained plaque produced by PV(M) (I) or an infected focus produced by PV-AB²⁴⁷⁰⁻²⁹⁵⁴ (J) after 48 h of infection.

FIG. 4. Codon deoptimization leads to a reduction of specific infectivity. (A) Agarose gel electrophoresis of virion genomic RNA isolated from purified virus particles of PV(M) (lane 1), PV-AB⁷⁵⁵⁻¹⁵¹³ (lane 2), and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ (lane 3). (B) Silver-stained SDS-PAGE protein gel of purified PV(M) (lane 1), PV-AB⁷⁵⁵⁻¹⁵¹³ (lane 2), and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ (lane 3) virus particles. The three larger of the four capsid proteins (VP1, VP2, and VP3) are shown, demonstrating the purity and relative amounts of virus preparations. (C) Development of a virus capture ELISA using a poliovirus receptor-alkaline phosphatase (CD155-AP) fusion protein probe. Virus-specific antibodies were used to coat ELISA plates, and samples containing an unknown virus concentration were applied followed by detection with CD155-AP. Virus concentrations were calculated using a standard curve prepared in parallel with known amounts of purified wt virus (E). (D) The amounts of purified virus and extracted virion RNA were spectrophotometrically quantified, and the number of particles or genome equivalents (1 genome=1 virion) was calculated. In addition, virion concentrations were determined by ELISA. The infectious titer of each virus was determined by plaque/infected-focus assay, and the specific infectivity was calculated as PFU/particle or FFU/particle.

FIG. 5. In vitro translation of codon-deoptimized and wild type viruses. The PV-AB phenotype is determined at the level of genome translation. (A) A standard in vitro translation in HeLa S10 extract, in the presence of exogenously added amino acids and tRNAs reveals no differences in translation capacities of codon-deoptimized genomes compared to the PV(M) wt. Shown is an autoradiograph of [³⁵S]methionine-labeled translation products resolved on a 12.5% SDS-PAGE gel. The identity of an aberrant band (*) is not known. (B) In vitro translation in nondialyzed HeLa S10 extract without the addition of exogenous amino acids and tRNA and in the presence of competing cellular mRNAs uncovers a defect in translation capacities of codon-deoptimized PV genomes. Shown is a Western blot of poliovirus

2C reactive translation products (2C^{ATPase}, 2BC, and P2) resolved on a 10% SDS-PAGE gel. The relative amounts of the 2BC translation products are expressed below each lane as percentages of the wt band.

FIG. 6. Analysis of in vivo translation using dicistronic reporter replicons confirms the detrimental effect of codon deoptimization on PV translation. (A) Schematic of dicistronic replicons. Various P1 capsid coding sequences were inserted upstream of the firefly luciferase gene (F-Luc). (B) Replicon RNAs were transfected into HeLa cells and incubated for 7 h in the presence of 2 mM guanidine-hydrochloride to block RNA replication. The relative rate of translation through the P1 region was inversely proportional to the extent of codon deoptimization. While the capsid coding sequences of two viable virus constructs, PV-AB²⁴⁷⁰⁻²⁹⁵⁴ and PV-AB²⁹⁵⁴⁻³³⁸⁶, allow between 60 and 80% of wt translation, translation efficiency below 20% is associated with the lethal phenotypes observed with the PV-AB, PV-AB²⁴⁷⁰⁻³³⁸⁶, and PV-AB¹⁵¹³⁻²⁴⁷⁰ genomes. Values represent the average of 6 assays from 3 independent experiments.

FIG. 7. Determining codon pair bias of human and viral ORFs. Dots represent the average codon-pair score per codon pair for one ORF plotted against its length. Codon pair bias (CPB) was calculated for 14,795 annotated human genes. Under-represented codon pairs yield negative scores. (C) CPB is plotted for various poliovirus P1 constructs, represented by symbols with arrows. The figure illustrates that the bulk of human genes clusters around 0.1. CPB is shown for PV(M)-wt (labeled "WT") (-0.02), customized synthetic poliovirus capsids PV-Max (+0.25), PV-Min (-0.48), and PV(M)-wt:PV-Min chimera capsids PV-Min⁷⁵⁵⁻²⁴⁷⁰ ("PV-MinXY") (-0.31) and PV-Min²⁴⁷⁰⁻³³⁸⁶ ("PV-MinZ") (-0.20). Viruses PV-SD and PV-AB are the result of altered codon bias, but not altered codon pair bias.

FIG. 8. Characteristics of codon-pair deoptimized polio. One-step growth kinetics reveals PFU production for PV-Min⁷⁵⁵⁻²⁴⁷⁰ and PV-Min²⁴⁷⁰⁻³³⁸⁶ that is reduced on the order of 2.5 orders of magnitude by comparison to PV(M)-wt. However, all viruses produce a similar number of viral particles (not shown in this Figure). As a result the PFU/particle ratio is reduced, similar to codon deoptimized viruses PV-AB⁷⁵⁵⁻¹⁵¹³ and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ (see FIG. 3B) (PFU is "Plaque Forming Unit").

FIG. 9. Assembly of chimeric viral genomes. To "scan" through a target genome (red) small segments are amplified or synthesized and introduced into the wt genome (black) by overlapping PCR.

FIG. 10. The eight-plasmid pol I-pol II system for the generation of influenza A virus. Eight expression plasmids containing the eight viral cDNAs inserted between the human pol I promoter and the pol II promoter are transfected into eukaryotic cells. Because each plasmid contains two different promoters, both cellular pol I and pol II will transcribe the plasmid template, presumably in different nuclear compartments, which will result in the synthesis of viral mRNAs and vRNAs. After synthesis of the viral polymerase complex proteins (PB1, PB2, PA, nucleoproteins), the viral replication cycle is initiated. Ultimately, the assembly of all viral molecules directly (pol II transcription) or indirectly (pol I transcription and viral replication) derived from the cellular transcription and translation machinery results in the interaction of all synthesized molecules (vRNPs and the structural proteins HA, NA, M1, M2,

NS2/NEP) to generate infectious influenza A virus. (Reproduced from Neumann et al., 2000.) (Note: there are other ways of synthesizing influenza de novo).

FIG. 11. Poliovirus Genome and Synthetic Viral Constructs. The poliovirus genome and open reading frames of chimeric virus constructs. Top, a schematic of the full-length PV(M)-wt genomic RNA. Below, the open reading frames of PV(M)-wt, the CPB customized synthetic viruses PV-Max, PV-Min, and the PV(M)-wt:PV-Min chimera viruses. Black corresponds to PV(M)-wt sequence, Gray to PV-Min synthetic sequence, and Thatched to PV-Max. The viral constructs highlighted, PV-Min⁷⁵⁵⁻²⁴⁷⁰ (PV-MinXY) and PV-Min²⁴⁷⁰⁻³³⁸⁵ (PV-MinZ), were further characterized due to a markedly attenuated phenotype.

FIG. 12. On-Step growth curves display similar kinetics yielding a similar quantity of particles with decreased infectivity. (A) An MOI of 2 was used to infect a monolayer of HeLa R19 cells, the PFU at the given time points (0, 2, 4, 7, 10, 24, 48 hrs) was measured by plaque assay. Corresponding symbols: (□) PV(M)-wt, (●) PV-Max, (◇) PV-Min⁷⁵⁵⁻¹⁵¹³, (x) PV-Min¹⁵¹³⁻²⁴⁷⁰, (◆) PV-MinXY, (Δ) PV-MinZ. (B) Displays the conversion of the calculated PFU/ml at each time point to particles/ml. This achieved by multiplying the PFU/ml by the respective viruses specific infectivity. Corresponding symbols as in (A)

FIG. 13. In vivo modulation of translation by alteration of CPB. (A) The dicistronic RNA construct used to quantify the in vivo effect CPB has on translation. The first cistron utilizes a hepatitis C virus (HCV) Internal Ribosome Entry Site (IRES) inducing the translation of *Renilla Luciferase* (R-Luc). This first cistron is the internal control used to normalize the amount of input RNA. The second cistron controlled by the PV(M)-wt IRES induces the translation of Firefly Luciferase (F-Luc). The region labeled "P1" in the construct was replaced by the cDNA of each respective viruses P1. (B) Each respective RNA construct was transfected, in the presence of 2 mM guanidine hydrochloride, into HeLa R19 cells and after 6 hours the R-Luc and F-Luc were measured. The F-Luc/R-Luc values were normalized relative to PV(M)-wt translation (100%).

FIG. 14. The heat inactivation profile of the synthetic viruses is unchanged. To rule out that large scale codon-pair bias modification alters the gross morphology of virions, as one might expect if capsid proteins were misfolded, the thermal stability of PVMiNXY and PV-MinZ was tested. An equal number of particles were incubated at 50° C. and the remaining infectivity quantified after given periods of time via plaque assay. If the capsids of the synthetic viruses were destabilized we would expect increased loss of viability at 50° C. in comparison to wt PV(M). This was not the case. The thermal inactivation kinetics of both synthetic viruses was identical to the wt. In contrast, the Sabin-1 virus carries numerous mutations in the genome region encoding the capsid, which, fittingly, rendered this virus less heat stable as compared to wt PV1(M).

FIG. 15. Neutralizing antibody titer following vaccination. A group of eight CD155 tg mice, seven of which completed the regimen, were each inoculated by intraperitoneal injection three times at weekly intervals with 10⁸ particles of PV-MinZ (❀) and PV-MinXY (◆) and the serum conversion was measured 10 days after the final vaccination. A horizontal lines across each data set marks the average neutralizing antibody titer for each virus construct. The anti-poliovirus antibody titer was measured via micro-neutralization assay. (*) No virus neutralization for mock-vaccinated animals was detected at the lowest tested 1:8.

FIG. 16. Influenza virus carrying codon pair-deoptimized NP segment. (A) A/PR8-NP^{Min} virus are viable and produce smaller plaques on MDCK cells compared to the A/PR8 wt. (B) A/PR8-NP^{Min} virus display delayed growth kinetics and final titers 3-5 fold below wild type A/PR8.

FIG. 17. Influenza virus carrying codon pair-deoptimized PB1 or HA and NP segments. (A) A/PR8-PB1^{Min-RR} and A/PR8-HA^{Min}/NP^{Min} virus are viable and produce smaller plaques on MDCK cells as compared to the A/PR8 wild type. (B) A/PR8-PB1^{Min-RR} and A/PR8-HA^{Min}/NP^{Min} virus display delayed growth kinetics and final titers about 10 fold below wild type A/PR8.

FIG. 18. Attenuation of A/PR8-NP^{Min} in BALB/c mouse model. (A) A/PR8-NP^{Min} virus has reduced pathogenicity compared to wild type A/PR8 virus as determined by weight loss upon vaccination. (B) All mice (eight of eight) vaccinated with A/PR8-NP^{Min} virus survived, whereas only 25% (two of eight) mice infected with A/PR8 were alive 13 days post vaccination. (C) Mice vaccinated with A/PR8-NP^{Min} virus are protected from challenge with 100×LD₅₀ of A/PR8 wild type virus.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to the production of attenuated viruses that may be used as vaccines to protect against viral infection and disease. Accordingly, the invention provides an attenuated virus, which comprises a modified viral genome containing nucleotide substitutions engineered in multiple locations in the genome, wherein the substitutions introduce a plurality of synonymous codons into the genome and/or a change of the order of existing codons for the same amino acid (change of codon pair utilization). In both cases, the original, wild-type amino acid sequences of the viral gene products are retained.

Most amino acids are encoded by more than one codon. See the genetic code in Table 1. For instance, alanine is encoded by GCU, GCC, GCA, and GCG. Three amino acids (Leu, Ser, and Arg) are encoded by six different codons, while only Trp and Met have unique codons. "Synonymous" codons are codons that encode the same amino acid. Thus, for example, CUU, CUC, CUA, CUG, UUA, and UUG are synonymous codons that code for Leu. Synonymous codons are not used with equal frequency. In general, the most frequently used codons in a particular organism are those for which the cognate tRNA is abundant, and the use of these codons enhances the rate and/or accuracy of protein translation. Conversely, tRNAs for the rarely used codons are found at relatively low levels, and the use of rare codons is thought to reduce translation rate and/or accuracy. Thus, to replace a given codon in a nucleic acid by a synonymous but less frequently used codon is to substitute a "deoptimized" codon into the nucleic acid.

TABLE 1

Genetic Code						
	U	C	A	G		
	U	Phe	Ser	Tyr	Cys	U
60		Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	STOP	A
	Leu	Ser	STOP	STOP	Trp	G
	C	Leu	Pro	His	Arg	U
65		Leu	Pro	His	Arg	C

TABLE 1-continued

Genetic Code				
	U	C	A	G
	Leu	Pro	Gln	Arg
	Leu	Pro	Gln	Arg
A	Ile	Thr	Asn	Ser
	Ile	Thr	Asn	Ser
	Ile	Thr	Lys	Arg
	Met	Thr	Lys	Arg
G	Val	Ala	Asp	Gly
	Val	Ala	Asp	Gly
	Val	Ala	Glu	Gly
	Val	Ala	Glu	Gly

^a The first nucleotide in each codon encoding a particular amino acid is shown in the left-most column; the second nucleotide is shown in the top row; and the third nucleotide is shown in the right-most column.

In addition, a given organism has a preference for the nearest codon neighbor of a given codon A, referred to a bias in codon pair utilization. A change of codon pair bias, without changing the existing codons, can influence the rate of protein synthesis and production of a protein.

In various embodiments of the present invention, the virus is a DNA, RNA, double-stranded, or single-stranded virus. In further embodiments, the virus infects an animal or a plant. In preferred embodiments, the animal is a human. A large number of animal viruses are well known to cause diseases (see below). Certain medically important viruses, such as those causing rabies, severe acute respiratory syndrome (SARS), and avian flu, can also spread to humans from their normal non-human hosts.

Viruses also constitute a major group of plant pathogens, and research is ongoing to develop viral vectors for producing transgenic plants. The advantages of such vectors include the ease of transforming plants, the ability to transform mature plants which obviates the need for regeneration of a transgenic plant from a single transformed cell, and high levels of expression of foreign genes from the multiple copies of virus per cell. However, one of the main disadvantages of these vectors is that it has not been possible to separate essential viral replicative functions from pathogenic determinants of the virus. The SAVE strategy disclosed herein may afford a means of engineering non-pathogenic viral vectors for plant transformation.

Major Viral Pathogens in Humans

Viral pathogens are the causative agents of many diseases in humans and other animals. Well known examples of viral diseases in humans include the common cold (caused by human rhinoviruses, HRV), influenza (influenza virus), chickenpox (varicella-zoster virus), measles (a paramyxovirus), mumps (a paramyxovirus), poliomyelitis (poliovirus, PV), rabies (Lyssavirus), cold sores (Herpes Simplex Virus [HSV] Type 1), and genital herpes (HSV Type 2). Prior to the introduction of vaccination programs for children, many of these were common childhood diseases worldwide, and are still a significant threat to health in some developing countries. Viral diseases also include more serious diseases such as acquired immunodeficiency syndrome (AIDS) caused by Human Immunodeficiency Virus (HIV), severe acute respiratory syndrome (SARS) caused by SARS coronavirus, avian flu (H5N1 subtype of influenza A virus), Ebola (ebolavirus), Marburg haemorrhagic fever (Marburg virus), dengue fever (Flavivirus serotypes), West Nile encephalitis (a flavivirus), infectious mononucleosis (Epstein-Barr virus, EBV), hepatitis (Hepatitis C Virus, HCV;

hepatitis B virus, HBV), and yellow fever (flavivirus). Certain types of cancer can also be caused by viruses. For example, although most infections by human papillomavirus (HPV) are benign, HPV has been found to be associated with cervical cancer, and Kaposi's sarcoma (KS), a tumor prevalent in AIDS patients, is caused by Kaposi's sarcoma-associated herpesvirus (KSHV).

Because viruses reside within cells and use the machinery of the host cell to reproduce, they are difficult to eliminate without killing the host cell. The most effective approach to counter viral diseases has been the vaccination of subjects at risk of infection in order to provide resistance to infection. For some diseases (e.g., chickenpox, measles, mumps, yellow fever), effective vaccines are available. However, there is a pressing need to develop vaccines for many other viral diseases. The SAVE (Synthetic Attenuated Virus Engineering) approach to making vaccines described herein is in principle applicable to all viruses for which a reverse genetics system (see below) is available. This approach is exemplified herein by focusing on the application of SAVE to develop attenuated virus vaccines for poliomyelitis, the common cold, and influenza.

Any virus can be attenuated by the methods disclosed herein. The virus can be a dsDNA virus (e.g. Adenoviruses, Herpesviruses, Poxviruses), a single stranded "plus" sense DNA virus (e.g., Parvoviruses) a double stranded RNA virus (e.g., Reoviruses), a single stranded+sense RNA virus (e.g. Picornaviruses, Togaviruses), a single stranded "minus" sense RNA virus (e.g. Orthomyxoviruses, Rhabdoviruses), a single stranded+sense RNA virus with a DNA intermediate (e.g. Retroviruses), or a double stranded reverse transcribing virus (e.g. Hepadnaviruses). In certain non-limiting embodiments of the present invention, the virus is poliovirus (PV), rhinovirus, influenza virus including avian flu (e.g. H5N1 subtype of influenza A virus), severe acute respiratory syndrome (SARS) coronavirus, Human Immunodeficiency Virus (HIV), Hepatitis B Virus (HBV), Hepatitis C Virus (HCV), infectious bronchitis virus, ebolavirus, Marburg virus, dengue fever virus (Flavivirus serotypes), West Nile disease virus, Epstein-Barr virus (EBV), yellow fever virus, Ebola (ebolavirus), chickenpox (varicella-zoster virus), measles (a paramyxovirus), mumps (a paramyxovirus), rabies (Lyssavirus), human papillomavirus, Kaposi's sarcoma-associated herpesvirus, Herpes Simplex Virus (HSV Type 1), or genital herpes (HSV Type 2).

The term "parent" virus or "parent" protein encoding sequence is used herein to refer to viral genomes and protein encoding sequences from which new sequences, which may be more or less attenuated, are derived. Parent viruses and sequences are usually "wild type" or "naturally occurring" prototypes or isolates of variants for which it is desired to obtain a more highly attenuated virus. However, parent viruses also include mutants specifically created or selected in the laboratory on the basis of real or perceived desirable properties. Accordingly, parent viruses that are candidates for attenuation include mutants of wild type or naturally occurring viruses that have deletions, insertions, amino acid substitutions and the like, and also include mutants which have codon substitutions. In one embodiment, such a parent sequence differs from a natural isolate by about 30 amino acids or fewer. In another embodiment, the parent sequence differs from a natural isolate by about 20 amino acids or fewer. In yet another embodiment, the parent sequence differs from a natural isolate by about 10 amino acids or fewer.

The attenuated PV may be derived from poliovirus type 1 (Mahoney; "PV(M)"), poliovirus type 2 (Lansing), poliovi-

rus type 3 (Leon), monovalent oral poliovirus vaccine (OPV) virus, or trivalent OPV virus. In certain embodiments, the poliovirus is PV-AB having the genomic sequence set forth in SEQ ID NO:2, or PV-AB⁷⁵⁵⁻¹⁵¹³, PV-AB⁷⁵⁵⁻²⁴⁷⁰, PV-AB¹⁵¹³⁻³³⁸⁶, PV-AB²⁴⁷⁰⁻³³⁸⁶, PV-AB¹⁵¹³⁻²⁴⁷⁰, PV-AB²⁴⁷⁰⁻²⁹⁵⁴, or PV-AB²⁹⁵⁴⁻³³⁸⁶. The nomenclature reflects a PV(M) genome in which portions of the genome, are substituted with nucleotides of PV-AB. The superscript provides the nucleotide numbers of PV-AB that are substituted.

In various embodiments, the attenuated rhinovirus is a human rhinovirus (HRV) derived from HRV2, HRV14, *Human rhinovirus 10*; *Human rhinovirus 100*; *Human rhinovirus 11*; *Human rhinovirus 12*; *Human rhinovirus 13*; *Human rhinovirus 15*; *Human rhinovirus 16*; *Human rhinovirus 18*; *Human rhinovirus 19*; *Human rhinovirus 1A*; *Human rhinovirus 1B*; *Human rhinovirus 2*; *Human rhinovirus 20*; *Human rhinovirus 21*; *Human rhinovirus 22*; *Human rhinovirus 23*; *Human rhinovirus 24*; *Human rhinovirus 25*; *Human rhinovirus 28*; *Human rhinovirus 29*; *Human rhinovirus 30*; *Human rhinovirus 31*; *Human rhinovirus 32*; *Human rhinovirus 33*; *Human rhinovirus 34*; *Human rhinovirus 36*; *Human rhinovirus 38*; *Human rhinovirus 39*; *Human rhinovirus 40*; *Human rhino virus 41*; *Human rhinovirus 43*; *Human rhinovirus 44*; *Human rhinovirus 45*; *Human rhinovirus 46*; *Human rhinovirus 47*; *Human rhinovirus 49*; *Human rhinovirus 50*; *Human rhinovirus 51*; *Human rhinovirus 53*; *Human rhinovirus 54*; *Human rhinovirus 55*; *Human rhinovirus 56*; *Human rhinovirus 57*; *Human rhinovirus 58*; *Human rhinovirus 59*; *Human rhinovirus 60*; *Human rhinovirus 61*; *Human rhinovirus 62*; *Human rhinovirus 63*; *Human rhinovirus 64*; *Human rhinovirus 65*; *Human rhinovirus 66*; *Human rhinovirus 67*; *Human rhinovirus 68*; *Human rhinovirus 7*; *Human rhinovirus 71*; *Human rhinovirus 73*; *Human rhinovirus 74*; *Human rhinovirus 75*; *Human rhinovirus 76*; *Human rhinovirus 77*; *Human rhinovirus 78*; *Human rhinovirus 8*; *Human rhinovirus 80*; *Human rhinovirus 81*; *Human rhinovirus 82*; *Human rhinovirus 85*; *Human rhinovirus 88*; *Human rhinovirus 89*; *Human rhinovirus 9*; *Human rhinovirus 90*; *Human rhinovirus 94*; *Human rhinovirus 95*; *Human rhinovirus 96*; *Human rhinovirus 98*; *Human rhinovirus 14*; *Human rhinovirus 17*; *Human rhinovirus 26*; *Human rhinovirus 27*; *Human rhinovirus 3*; *Human rhinovirus 8001* Finland Nov1995; *Human rhinovirus 35*; *Human rhinovirus 37*; +*Human rhinovirus 6253* Finland Sep1994; *Human rhinovirus 9166* Finland Sep1995; *Human rhinovirus 4*; *Human rhinovirus 42*; *Human rhinovirus 48*; *Human rhinovirus 9864* Finland Sep1996; *Human rhinovirus 5*; *Human rhinovirus 52*; *Human rhinovirus 6*; *Human rhinovirus 7425* Finland Dec1995; *Human rhinovirus 69*; *Human rhinovirus 5928* Finland May1995; *Human rhinovirus 70*; *Human rhinovirus 72*; *Human rhinovirus 79*; *Human rhinovirus 83*; *Human rhinovirus 84*; *Human rhinovirus 8317* Finland Aug1996; *Human rhinovirus 86*; *Human rhinovirus 91*; *Human rhinovirus 7851* Finland Sep1996; *Human rhinovirus 92*; *Human rhinovirus 93*; *Human rhinovirus 97*; *Human rhinovirus 99*; Antwerp rhinovirus 98/99; *Human rhinovirus 263* Berlin 2004; *Human rhinovirus 3083/rhino/Hyogo/2005*; *Human rhinovirus NY-003*; *Human rhinovirus NY-028*; *Human rhinovirus NY-041*; *Human rhinovirus NY-042*; *Human rhinovirus NY-060*; *Human rhinovirus NY-063*; *Human rhinovirus NY-074*; *Human rhinovirus NY-1085*; *Human rhinovirus strain Hanks*; Untyped *human rhinovirus OK88-8162*; *Human enterovirus* sp. ex *Amblyomma americanum*; *Human rhinovirus* sp. or *Human rhinovirus UC*.

In other embodiments, the attenuated influenza virus is derived from influenza virus A, influenza virus B, or influenza virus C. In further embodiments, the influenza virus A belongs to but is not limited to subtype H10N7, H10N1, H10N2, H10N3, H10N4, H10N5, H10N6, H10N7, H10N8, H10N9, H11N1, H11N2, H11N3, H11N4, H11N6, H11N8, H11N9, H12N1, H12N2, H12N4, H12N5, H12N6, H12N8, H12N9, H13N2, H13N3, H13N6, H13N9, H14N5, H14N6, H15N2, H15N8, H15N9, H16N3, H1N1, H1N2, H1N3, H1N5, H1N6, H1N8, H1N9, H2N1, H2N2, H2N3, H2N4, H2N5, H2N6, H2N7, H2N8, H2N9, H3N1, H3N2, H3N3, H3N4, H3N5, H3N6, H3N8, H3N9, H4N1, H4N2, H4N3, H4N4, H4N5, H4N6, H4N7, H4N8, H4N9, H5N1, H5N2, H5N3, H5N4, H5N6, H5N7, H5N8, H5N9, H6N1, H6N2, H6N3, H6N4, H6N5, H6N6, H6N7, H6N8, H6N9, H7N1, H7N2, H7N3, H7N4, H7N5, H7N7, H7N8, H7N9, H8N2, H8N4, H8N5, H9N1, H9N2, H9N3, H9N4, H9N5, H9N6, H9N7, H9N8, H9N9 and unidentified subtypes. In further embodiments, the influenza virus B belongs to but is not limited to subtype Influenza B virus (B/Aichi/186/2005), Influenza B virus (B/Aichi/5/88), Influenza B virus (B/Akita/27/2001), Influenza B virus (B/Akita/5/2001), Influenza B virus (B/Alabama/1/2006), Influenza B virus (B/Alabama/2/2005), Influenza B virus (B/Alaska/03/1992), Influenza B virus (B/Alaska/12/1996), Influenza B virus (B/Alaska/16/2000), Influenza B virus (B/Alaska/16/2003), Influenza B virus (B/Alaska/1777/2005), Influenza B virus (B/Alaska/2/2004), Influenza B virus (B/Alaska/6/2005), Influenza B virus (B/Ann Arbor/1/1986), Influenza B virus (B/Ann Arbor/1994), Influenza B virus (B/Argentina/132/2001), Influenza B virus (B/Argentina/3640/1999), Influenza B virus (B/Argentina/69/2001), Influenza B virus (B/Arizona/1/2005), Influenza B virus (B/Arizona/12/2003), Influenza B virus (B/Arizona/13/2003), Influenza B virus (B/Arizona/135/2005), Influenza B virus (B/Arizona/14/2001), Influenza B virus (B/Arizona/14/2005), Influenza B virus (B/Arizona/140/2005), Influenza B virus (B/Arizona/148/2005), Influenza B virus (B/Arizona/15/2005), Influenza B virus (B/Arizona/16/2005), Influenza B virus (B/Arizona/162/2005), Influenza B virus (B/Arizona/163/2005), Influenza B virus (B/Arizona/164/2005), Influenza B virus (B/Arizona/2/2000), Influenza B virus (B/Arizona/2/2005), Influenza B virus (B/Arizona/2e/2006), Influenza B virus (B/Arizona/3/2006), Influenza B virus (B/Arizona/4/2002), Influenza B virus (B/Arizona/4/2006), Influenza B virus (B/Arizona/48/2005), Influenza B virus (B/Arizona/5/2000), Influenza B virus (B/Arizona/59/2005), Influenza B virus (B/Arizona/7/2000), Influenza B virus (B/Auckland/01/2000), Influenza B virus (B/Bangkok/141/1994), Influenza B virus (B/Bangkok/143/1994), Influenza B virus (B/Bangkok/153/1990), Influenza B virus (B/Bangkok/163/1990), Influenza B virus (B/Bangkok/163/90), Influenza B virus (B/Bangkok/34/99), Influenza B virus (B/Bangkok/460/03), Influenza B virus (B/Bangkok/54/99), Influenza B virus (B/Barcelona/215/03), Influenza B virus (B/Beijing/15/84), Influenza B virus (B/Beijing/184/93), Influenza B virus (B/Beijing/243/97), Influenza B virus (B/Beijing/43/75), Influenza B virus (B/Beijing/5/76), Influenza B virus (B/Beijing/76/98), Influenza B virus (B/Belgium/WV106/2002), Influenza B virus (B/Belgium/WV107/2002), Influenza B virus (B/Belgium/WV109/2002), Influenza B virus (B/Belgium/WV114/2002), Influenza B virus (B/Belgium/WV122/2002), Influenza B virus (B/Bonn/43), Influenza B virus (B/Brazil/017/00), Influenza B virus (B/Brazil/053/00), Influenza B virus (B/Brazil/055/00), Influenza B virus (B/Brazil/064/00), Influenza B virus (B/Brazil/074/00), Influenza B virus

(B/Brazil/079/00), Influenza B virus (B/Brazil/110/01), Influenza B virus (B/Brazil/952/2001), Influenza B virus (B/Brazil/975/2000), Influenza B virus (B/Brisbane/32/2002), Influenza B virus (B/Bucharest/311/1998), Influenza B virus (B/Bucharest/795/03), Influenza B virus (B/Buenos Aires/161/00), Influenza B virus (B/Buenos Aires/9/95), Influenza B virus (B/Buenos Aires/SW16/97), Influenza B virus (B/Buenos Aires/VL518/99), Influenza B virus (B/California/01/1995), Influenza B virus (B/California/02/1994), Influenza B virus (B/California/02/1995), Influenza B virus (B/California/1/2000), Influenza B virus (B/California/10/2000), Influenza B virus (B/California/11/2001), Influenza B virus (B/California/14/2005), Influenza B virus (B/California/2/2002), Influenza B virus (B/California/2/2003), Influenza B virus (B/California/3/2000), Influenza B virus (B/California/3/2004), Influenza B virus (B/California/6/2000), Influenza B virus (B/California/7/2005), Influenza B virus (B/Canada/16188/2000), Influenza B virus (B/Canada/464/2001), Influenza B virus (B/Canada/464/2002), Influenza B virus (B/Chaco/366/00), Influenza B virus (B/Chaco/R113/00), Influenza B virus (B/Chantaburi/218/2003), Influenza B virus (B/Cheju/303/03), Influenza B virus (B/Chiba/447/98), Influenza B virus (B/Chile/3162/2002), Influenza B virus (B/Chongqing/3/2000), Influenza B virus (B/clinical isolate SA1 Thailand/2002), Influenza B virus (B/clinical isolate SA10 Thailand/2002), Influenza B virus (B/clinical isolate SA100 Philippines/2002), Influenza B virus (B/clinical isolate SA101 Philippines/2002), Influenza B virus (B/clinical isolate SA102 Philippines/2002), Influenza B virus (B/clinical isolate SA103 Philippines/2002), Influenza B virus (B/clinical isolate SA104 Philippines/2002), Influenza B virus (B/clinical isolate SA105 Philippines/2002), Influenza B virus (B/clinical isolate SA106 Philippines/2002), Influenza B virus (B/clinical isolate SA107 Philippines/2002), Influenza B virus (B/clinical isolate SA108 Philippines/2002), Influenza B virus (B/clinical isolate SA109 Philippines/2002), Influenza B virus (B/clinical isolate SA11 Thailand/2002), Influenza B virus (B/clinical isolate SA110 Philippines/2002), Influenza B virus (B/clinical isolate SA112 Philippines/2002), Influenza B virus (B/clinical isolate SA113 Philippines/2002), Influenza B virus (B/clinical isolate SA114 Philippines/2002), Influenza B virus (B/clinical isolate SA115 Philippines/2002), Influenza B virus (B/clinical isolate SA116 Philippines/2002), Influenza B virus (B/clinical isolate SA12 Thailand/2002), Influenza B virus (B/clinical isolate SA13 Thailand/2002), Influenza B virus (B/clinical isolate SA14 Thailand/2002), Influenza B virus (B/clinical isolate SA15 Thailand/2002), Influenza B virus (B/clinical isolate SA16 Thailand/2002), Influenza B virus (B/clinical isolate SA17 Thailand/2002), Influenza B virus (B/clinical isolate SA18 Thailand/2002), Influenza B virus (B/clinical isolate SA19 Thailand/2002), Influenza B virus (B/clinical isolate SA2 Thailand/2002), Influenza B virus (B/clinical isolate SA20 Thailand/2002), Influenza B virus (B/clinical isolate SA21 Thailand/2002), Influenza B virus (B/clinical isolate SA22 Thailand/2002), Influenza B virus (B/clinical isolate SA23 Thailand/2002), Influenza B virus (B/clinical isolate SA24 Thailand/2002), Influenza B virus (B/clinical isolate SA25 Thailand/2002), Influenza B virus (B/clinical isolate SA26 Thailand/2002), Influenza B virus (B/clinical isolate SA27 Thailand/2002), Influenza B virus (B/clinical isolate SA28 Thailand/2002), Influenza B virus (B/clinical isolate SA29 Thailand/2002), Influenza B virus (B/clinical isolate SA3 Thailand/2002), Influenza B virus (B/clinical isolate SA30 Thailand/2002), Influenza B virus (B/clinical isolate SA31 Thailand/2002), Influenza B virus (B/clinical isolate SA32 Thailand/2002).

B virus (B/CNIC/27/2001), Influenza B virus (B/Colorado/04/2004), Influenza B virus (B/Colorado/11e/2004), Influenza B virus (B/Colorado/12e/2005), Influenza B virus (B/Colorado/13/2004), Influenza B virus (B/Colorado/13e/2004), Influenza B virus (B/Colorado/15/2004), Influenza B virus (B/Colorado/16e/2004), Influenza B virus (B/Colorado/17e/2004), Influenza B virus (B/Colorado/2/2004), Influenza B virus (B/Colorado/2597/2004), Influenza B virus (B/Colorado/4e/2004), Influenza B virus (B/Colorado/5/2004), Influenza B virus (B/Connecticut/02/1995), Influenza B virus (B/Connecticut/07/1993), Influenza B virus (B/Cordoba/2979/1991), Influenza B virus (B/Cordoba/VA418/99), Influenza B virus (B/Czechoslovakia/16/89), Influenza B virus (B/Czechoslovakia/69/1990), Influenza B virus (B/Czechoslovakia/69/90), Influenza B virus (B/Daeku/10/97), Influenza B virus (B/Daeku/45/97), Influenza B virus (B/Daeku/47/97), Influenza B virus (B/Daeku/9/97), Influenza B virus (B/Delaware/1/2006), Influenza B virus (B/Du/4/78), Influenza B virus (B/Durban/39/98), Influenza B virus (B/Durban/43/98), Influenza B virus (B/Durban/44/98), Influenza B virus (B/Durban/52/98), Influenza B virus (B/Durban/55/98), Influenza B virus (B/Durban/56/98), Influenza B virus (B/Egypt/2040/2004), Influenza B virus (B/England/1716/2005), Influenza B virus (B/England/2054/2005), Influenza B virus (B/England/23/04), Influenza B virus (B/EspiritoSanto/55/01), Influenza B virus (B/EspiritoSanto/79/99), Influenza B virus (B/Finland/154/2002), Influenza B virus (B/Finland/159/2002), Influenza B virus (B/Finland/160/2002), Influenza B virus (B/Finland/161/2002), Influenza B virus (B/Finland/162/03), Influenza B virus (B/Finland/162/2002), Influenza B virus (B/Finland/162/91), Influenza B virus (B/Finland/164/2003), Influenza B virus (B/Finland/172/91), Influenza B virus (B/Finland/173/2003), Influenza B virus (B/Finland/176/2003), Influenza B virus (B/Finland/184/91), Influenza B virus (B/Finland/188/2003), Influenza B virus (B/Finland/190/2003), Influenza B virus (B/Finland/191/2003), Influenza B virus (B/Finland/192/2003), Influenza B virus (B/Finland/193/2003), Influenza B virus (B/Finland/199/2003), Influenza B virus (B/Finland/202/2003), Influenza B virus (B/Finland/203/2003), Influenza B virus (B/Finland/204/2003), Influenza B virus (B/Finland/205/2003), Influenza B virus (B/Finland/206/2003), Influenza B virus (B/Finland/220/2003), Influenza B virus (B/Finland/223/2003), Influenza B virus (B/Finland/225/2003), Influenza B virus (B/Finland/227/2003), Influenza B virus (B/Finland/231/2003), Influenza B virus (B/Finland/235/2003), Influenza B virus (B/Finland/239/2003), Influenza B virus (B/Finland/244/2003), Influenza B virus (B/Finland/245/2003), Influenza B virus (B/Finland/254/2003), Influenza B virus (B/Finland/254/93), Influenza B virus (B/Finland/255/2003), Influenza B virus (B/Finland/268/93), Influenza B virus (B/Finland/270/2003), Influenza B virus (B/Finland/275/2003), Influenza B virus (B/Finland/767/2000), Influenza B virus (B/Finland/84/2002), Influenza B virus (B/Finland/886/2001), Influenza B virus (B/Finland/WV4/2002), Influenza B virus (B/Finland/WV5/2002), Influenza B virus (B/Florida/02/1998), Influenza B virus (B/Florida/02/2006), Influenza B virus (B/Florida/1/2000), Influenza B virus (B/Florida/1/2004), Influenza B virus (B/Florida/2/2004), Influenza B virus (B/Florida/2/2005), Influenza B virus (B/Florida/2/2006), Influenza B virus (B/Florida/7e/2004), Influenza B virus (B/Fujian/36/82), Influenza B virus (B/Geneva/5079/03), Influenza B virus (B/Genoa/11/02), Influenza B virus (B/Genoa/2/02), Influenza B virus (B/Genoa/21/02), Influenza B virus (B/Genoa/33/02), Influenza B virus (B/Genoa/

41/02), Influenza B virus (B/Genoa/52/02), Influenza B virus (B/Genoa/55/02), Influenza B virus (B/Genoa/56/02), Influenza B virus (B/Genoa/7/02), Influenza B virus (B/Genoa/8/02), Influenza B virus (B/Genoa12/02), Influenza B virus (B/Genoa3/02), Influenza B virus (B/Genoa48/02), Influenza B virus (B/Genoa49/02), Influenza B virus (B/Genoa5/02), Influenza B virus (B/Genoa53/02), Influenza B virus (B/Genoa6/02), Influenza B virus (B/Genoa65/02), Influenza B virus (B/Genova/1294/03), Influenza B virus (B/Genova/1603/03), Influenza B virus (B/Genova/2/02), Influenza B virus (B/Genova/20/02), Influenza B virus (B/Genova/2059/03), Influenza B virus (B/Genova/26/02), Influenza B virus (B/Genova/30/02), Influenza B virus (B/Genova/54/02), Influenza B virus (B/Genova/55/02), Influenza B virus (B/Georgia/02/1998), Influenza B virus (B/Georgia/04/1998), Influenza B virus (B/Georgia/09/2005), Influenza B virus (B/Georgia/1/2000), Influenza B virus (B/Georgia/1/2005), Influenza B virus (B/Georgia/2/2005), Influenza B virus (B/Georgia/9/2005), Influenza B virus (B/Guangdong/05/94), Influenza B virus (B/Guangdong/08/93), Influenza B virus (B/Guangdong/5/94), Influenza B virus (B/Guangdong/55/89), Influenza B virus (B/Guangdong/8/93), Influenza B virus (B/Guangzhou/7/97), Influenza B virus (B/Guangzhou/86/92), Influenza B virus (B/Guangzhou/87/92), Influenza B virus (B/Gyeonggi/592/2005), Influenza B virus (B/Hannover/2/90), Influenza B virus (B/Harbin/07/94), Influenza B virus (B/Hawaii/1/2003), Influenza B virus (B/Hawaii/10/2001), Influenza B virus (B/Hawaii/10/2004), Influenza B virus (B/Hawaii/11/2004), Influenza B virus (B/Hawaii/11e/2004), Influenza B virus (B/Hawaii/11e/2005), Influenza B virus (B/Hawaii/12e/2005), Influenza B virus (B/Hawaii/13/2004), Influenza B virus (B/Hawaii/13e/2004), Influenza B virus (B/Hawaii/17/2001), Influenza B virus (B/Hawaii/18e/2004), Influenza B virus (B/Hawaii/1990/2004), Influenza B virus (B/Hawaii/1993/2004), Influenza B virus (B/Hawaii/19e/2004), Influenza B virus (B/Hawaii/2/2000), Influenza B virus (B/Hawaii/2/2003), Influenza B virus (B/Hawaii/20e/2004), Influenza B virus (B/Hawaii/21/2004), Influenza B virus (B/Hawaii/26/2001), Influenza B virus (B/Hawaii/31e/2004), Influenza B virus (B/Hawaii/32e/2004), Influenza B virus (B/Hawaii/33e/2004), Influenza B virus (B/Hawaii/35/2001), Influenza B virus (B/Hawaii/36/2001), Influenza B virus (B/Hawaii/37/2001), Influenza B virus (B/Hawaii/4/2006), Influenza B virus (B/Hawaii/43/2001), Influenza B virus (B/Hawaii/44/2001), Influenza B virus (B/Hawaii/9/2001), Influenza B virus (B/Hebei/19/94), Influenza B virus (B/Hebei/3/94), Influenza B virus (B/Hebei/4/95), Influenza B virus (B/Henan/22/97), Influenza B virus (B/Hiroshima/23/2001), Influenza B virus (B/Hong Kong/02/1993), Influenza B virus (B/Hong Kong/03/1992), Influenza B virus (B/Hong Kong/05/1972), Influenza B virus (B/Hong Kong/06/2001), Influenza B virus (B/Hong Kong/110/99), Influenza B virus (B/Hong Kong/1115/2002), Influenza B virus (B/Hong Kong/112/2001), Influenza B virus (B/Hong Kong/123/2001), Influenza B virus (B/Hong Kong/1351/02), Influenza B virus (B/Hong Kong/1351/2002), Influenza B virus (B/Hong Kong/1434/2002), Influenza B virus (B/Hong Kong/147/99), Influenza B virus (B/Hong Kong/156/99), Influenza B virus (B/Hong Kong/157/99), Influenza B virus (B/Hong Kong/167/2002), Influenza B virus (B/Hong Kong/22/1989), Influenza B virus (B/Hong Kong/22/2001), Influenza B virus (B/Hong Kong/22/89), Influenza B virus (B/Hong Kong/28/2001), Influenza B virus (B/Hong Kong/293/02), Influenza B virus (B/Hong Kong/310/2004), Influenza B virus (B/Hong Kong/329/2001), Influenza B virus

(B/Hong Kong/330/2001 egg adapted), Influenza B virus (B/Hong Kong/330/2001), Influenza B virus (B/Hong Kong/330/2002), Influenza B virus (B/Hong Kong/335/2001), Influenza B virus (B/Hong Kong/336/2001), Influenza B virus (B/Hong Kong/497/2001), Influenza B virus (B/Hong Kong/542/2000), Influenza B virus (B/Hong Kong/548/2000), Influenza B virus (B/Hong Kong/553a/2003), Influenza B virus (B/Hong Kong/557/2000), Influenza B virus (B/Hong Kong/6/2001), Influenza B virus (B/Hong Kong/666/2001), Influenza B virus (B/Hong Kong/692/01), Influenza B virus (B/Hong Kong/70/1996), Influenza B virus (B/Hong Kong/8/1973), Influenza B virus (B/Hong Kong/9/89), Influenza B virus (B/Houston/1/91), Influenza B virus (B/Houston/1/92), Influenza B virus (B/Houston/1/96), Influenza B virus (B/Houston/2/93), Influenza B virus (B/Houston/2/96), Influenza B virus (B/Houston/B15/1999), Influenza B virus (B/Houston/B56/1997), Influenza B virus (B/Houston/B57/1997), Influenza B virus (B/Houston/B58/1997), Influenza B virus (B/Houston/B59/1997), Influenza B virus (B/Houston/B60/1997), Influenza B virus (B/Houston/B61/1997), Influenza B virus (B/Houston/B63/1997), Influenza B virus (B/Houston/B65/1998), Influenza B virus (B/Houston/B66/2000), Influenza B virus (B/Houston/B67/2000), Influenza B virus (B/Houston/B68/2000), Influenza B virus (B/Houston/B69/2002), Influenza B virus (B/Houston/B70/2002), Influenza B virus (B/Houston/B71/2002), Influenza B virus (B/Houston/B720/2004), Influenza B virus (B/Houston/B74/2002), Influenza B virus (B/Houston/B745/2005), Influenza B virus (B/Houston/B75/2002), Influenza B virus (B/Houston/B756/2005), Influenza B virus (B/Houston/B77/2002), Influenza B virus (B/Houston/B787/2005), Influenza B virus (B/Houston/B79/2003), Influenza B virus (B/Houston/B81/2003), Influenza B virus (B/Houston/B84/2003), Influenza B virus (B/Houston/B846/2005), Influenza B virus (B/Houston/B850/2005), Influenza B virus (B/Houston/B86/2003), Influenza B virus (B/Houston/B87/2003), Influenza B virus (B/Houston/B88/2003), Influenza B virus (B/Hunan/4/72), Influenza B virus (B/Ibaraki/2/85), Influenza B virus (B/Idaho/1/2005), Influenza B virus (B/Illinois/1/2004), Influenza B virus (B/Illinois/13/2004), Influenza B virus (B/Illinois/13/2005), Influenza B virus (B/Illinois/13e/2005), Influenza B virus (B/Illinois/3/2001), Influenza B virus (B/Illinois/3/2005), Influenza B virus (B/Illinois/33/2005), Influenza B virus (B/Illinois/36/2005), Influenza B virus (B/Illinois/4/2005), Influenza B virus (B/Illinois/47/2005), Influenza B virus (B/Incheon/297/2005), Influenza B virus (B/India/3/89), Influenza B virus (B/India/7526/2001), Influenza B virus (B/India/7569/2001), Influenza B virus (B/India/7600/2001), Influenza B virus (B/India/7605/2001), Influenza B virus (B/India/77276/2001), Influenza B virus (B/Indiana/01/1995), Influenza B virus (B/Indiana/3/2006), Influenza B virus (B/Indiana/5/2006), Influenza B virus (B/Iowa/03/2002), Influenza B virus (B/Iowa/1/2001), Influenza B virus (B/Iowa/1/2005), Influenza B virus (B/Israel/95/03), Influenza B virus (B/Israel/WV 124/2002), Influenza B virus (B/Israel/WV126/2002), Influenza B virus (B/Israel/WV133/2002), Influenza B virus (B/Israel/WV135/2002), Influenza B virus (B/Israel/WV137/2002), Influenza B virus (B/Israel/WV142/2002), Influenza B virus (B/Israel/WV143/2002), Influenza B virus (B/Israel/WV145/2002), Influenza B virus (B/Israel/WV146/2002), Influenza B virus (B/Israel/WV150/2002), Influenza B virus (B/Israel/WV153/2002), Influenza B virus (B/Israel/WV158/2002), Influenza B virus (B/Israel/WV161/2002), Influenza B virus (B/Israel/WV166/2002), Influenza B virus (B/Israel/WV169/2002), Influenza B virus (B/Israel/WV170/2002),

Influenza B virus (B/Israel/WV174/2002), Influenza B virus (B/Israel/WV183/2002), Influenza B virus (B/Israel/WV187/2002), Influenza B virus (B/Istanbul/CTF-132/05), Influenza B virus (B/Japan/1224/2005), Influenza B virus (B/Japan/1905/2005), Influenza B virus (B/Jiangsu/10/03), Influenza B virus (B/Jiangsu/10/2003 (recomb)), Influenza B virus (B/Jiangsu/10/2003), Influenza B virus (B/Jilin/20/2003), Influenza B virus (B/Johannesburg/05/1999), Influenza B virus (B/Johannesburg/06/1994), Influenza B virus (B/Johannesburg/1/99), Influenza B virus (B/Johannesburg/113/010), Influenza B virus (B/Johannesburg/116/01), Influenza B virus (B/Johannesburg/119/01), Influenza B virus (B/Johannesburg/123/01), Influenza B virus (B/Johannesburg/163/99), Influenza B virus (B/Johannesburg/187/99), Influenza B virus (B/Johannesburg/189/99), Influenza B virus (B/Johannesburg/2/99), Influenza B virus (B/Johannesburg/27/2005), Influenza B virus (B/Johannesburg/33/01), Influenza B virus (B/Johannesburg/34/01), Influenza B virus (B/Johannesburg/35/01), Influenza B virus (B/Johannesburg/36/01), Influenza B virus (B/Johannesburg/41/99), Influenza B virus (B/Johannesburg/5/99), Influenza B virus (B/Johannesburg/69/2001), Influenza B virus (B/Johannesburg/77/01), Influenza B virus (B/Johannesburg/94/99), Influenza B virus (B/Johannesburg/96/01), Influenza B virus (B/Kadoma/1076/99), Influenza B virus (B/Kadoma/122/99), Influenza B virus (B/Kadoma/122/99-V1), Influenza B virus (B/Kadoma/122/99-V10), Influenza B virus (B/Kadoma/122/99-V11), Influenza B virus (B/Kadoma/122/99-V2), Influenza B virus (B/Kadoma/122/99-V3), Influenza B virus (B/Kadoma/122/99-V4), Influenza B virus (B/Kadoma/122/99-V5), Influenza B virus (B/Kadoma/122/99-V6), Influenza B virus (B/Kadoma/122/99-V7), Influenza B virus (B/Kadoma/122/99-V8), Influenza B virus (B/Kadoma/122/99-V9), Influenza B virus (B/Kadoma/136/99), Influenza B virus (B/Kadoma/409/2000), Influenza B virus (B/Kadoma/506/99), Influenza B virus (B/kadoma/642/99), Influenza B virus (B/Kadoma/647/99), Influenza B virus (B/Kagoshima/15/94), Influenza B virus (B/Kanagawa/73), Influenza B virus (B/Kansas/1/2005), Influenza B virus (B/Kansas/22992/99), Influenza B virus (B/Kentucky/4/2005), Influenza B virus (B/Khazkov/224/91), Influenza B virus (B/Kisumu/2036/2006), Influenza B virus (B/Kisumu/2037/2006), Influenza B virus (B/Kisumu/2038/2006), Influenza B virus (B/Kisumu/2039/2006), Influenza B virus (B/Kisumu/2040/2006), Influenza B virus (B/Kisumu/7/2005), Influenza B virus (B/Kobe/1/2002), Influenza B virus (B/Kobe/1/2002-V1), Influenza B virus (B/Kobe/1/2003), Influenza B virus (B/Kobe/1/94), Influenza B virus (B/Kobe/2/2002), Influenza B virus (B/Kobe/2/2003), Influenza B virus (B/Kobe/25/2003), Influenza B virus (B/Kobe/26/2003), Influenza B virus (B/Kobe/28/2003), Influenza B virus (B/Kobe/3/2002), Influenza B virus (B/Kobe/3/2003), Influenza B virus (B/Kobe/4/2002), Influenza B virus (B/Kobe/4/2003), Influenza B virus (B/Kobe/5/2002), Influenza B virus (B/Kobe/6/2002), Influenza B virus (B/Kobe/64/2001), Influenza B virus (B/Kobe/65/2001), Influenza B virus (B/Kobe/69/2001), Influenza B virus (B/Kobe/7/2002), Influenza B virus (B/Kobe/79/2001), Influenza B virus (B/Kobe/83/2001), Influenza B virus (B/Kobe/87/2001), Influenza B virus (B/Kouchi/193/1999), Influenza B virus (B/Kouchi/193/99), Influenza B virus (B/Lazio/1/02), Influenza B virus (B/Lee/40), Influenza B virus (B/Leningrad/129/91), Influenza B virus (B/Leningrad/148/91), Influenza B virus (B/Lisbon/02/1994), Influenza B virus (B/Lissabon/2/90), Influenza B virus (B/Los Angeles/1/02), Influenza B virus (B/Lusaka/270/99), Influenza B virus (B/Lusaka/432/99), Influenza B

19

virus (B/Lyon/1271/96), Influenza B virus (B/Malaysia/83077/2001), Influenza B virus (B/Maputo/1/99), Influenza B virus (B/Maputo/2/99), Influenza B virus (B/Mar del Plata/595/99), Influenza B virus (B/Mar del Plata/VL373/99), Influenza B virus (B/Mar del Plata/VL385/99), Influenza B virus (B/Maryland/1/01), Influenza B virus (B/Maryland/1/2002), Influenza B virus (B/Maryland/2/2001), Influenza B virus (B/Maryland/7/2003), Influenza B virus (B/Massachusetts/1/2004), Influenza B virus (B/Massachusetts/2/2004), Influenza B virus (B/Massachusetts/3/2004), Influenza B virus (B/Massachusetts/4/2001), Influenza B virus (B/Massachusetts/5/2003), Influenza B virus (B/Memphis/1/01), Influenza B virus (B/Memphis/10/97), Influenza B virus (B/Memphis/11/2006), Influenza B virus (B/Memphis/12/2006), Influenza B virus (B/Memphis/12/97), Influenza B virus (B/Memphis/12/97-MA), Influenza B virus (B/Memphis/13/03), Influenza B virus (B/Memphis/18/95), Influenza B virus (B/Memphis/19/96), Influenza B virus (B/Memphis/20/96), Influenza B virus (B/Memphis/21/96), Influenza B virus (B/Memphis/28/96), Influenza B virus (B/Memphis/3/01), Influenza B virus (B/Memphis/3/89), Influenza B virus (B/Memphis/3/93), Influenza B virus (B/Memphis/4/93), Influenza B virus (B/Memphis/5/93), Influenza B virus (B/Memphis/7/03), Influenza B virus (B/Memphis/8/99), Influenza B virus (B/Mexico/84/2000), Influenza B virus (B/Michigan/04/2006), Influenza B virus (B/Michigan/)/2005), Influenza B virus (B/Michigan/)/2006), Influenza B virus (B/Michigan/2/2004), Influenza B virus (B/Michigan/20/2005), Influenza B virus (B/Michigan/22572/99), Influenza B virus (B/Michigan/22587/99), Influenza B virus (B/Michigan/22596/99), Influenza B virus (B/Michigan/22631/99), Influenza B virus (B/Michigan/22659/99), Influenza B virus (B/Michigan/22687/99), Influenza B virus (B/Michigan/22691/99), Influenza B virus (B/Michigan/22721/99), Influenza B virus (B/Michigan/22723/99), Influenza B virus (B/Michigan/2e/2006), Influenza B virus (B/Michigan/3/2004), Influenza B virus (B/Michigan/4/2006), Influenza B virus (B/Michigan/e3/2006), Influenza B virus (B/micona/1/1989), Influenza B virus (B/Mie/01/1993), Influenza B virus (B/Mie/1/93), Influenza B virus (B/Milano/1/01), Influenza B virus (B/Milano/1/02), Influenza B virus (B/Milano/5/02), Influenza B virus (B/Milano/6/02), Influenza B virus (B/Milano/66/04), Influenza B virus (B/Milano/7/02), Influenza B virus (B/Minnesota/1/1985), Influenza B virus (B/Minnesota/14/2001), Influenza B virus (B/Minnesota/2/2001), Influenza B virus (B/Minsk/318/90), Influenza B virus (B/Mississippi/)/2001), Influenza B virus (B/Mississippi/2/2005), Influenza B virus (B/Mississippi/3/2001), Influenza B virus (B/Mississippi/3/2005), Influenza B virus (B/Mississippi/4/2003), Influenza B virus (B/Mississippi/4e/2005), Influenza B virus (B/Missouri/1/2006), Influenza B virus (B/Missouri/11/2003), Influenza B virus (B/Missouri/2/2005), Influenza B virus (B/Missouri/20/2003), Influenza B virus (B/Missouri/6/2005), Influenza B virus (B/Montana/1/2003), Influenza B virus (B/Montana/1/2006), Influenza B virus (B/Montana/1e/2004), Influenza B virus (B/Moscow/16/2002), Influenza B virus (B/Moscow/3/03), Influenza B virus (B/Nagoya/20/99), Influenza B virus (B/Nairobi/2032/2006), Influenza B virus (B/Nairobi/2033/2006), Influenza B virus (B/Nairobi/2034/2006), Influenza B virus (B/Nairobi/2035/2006), Influenza B virus (B/Nairobi/351/2005), Influenza B virus (B/Nairobi/670/2005), Influenza B virus (B/Nanchang/1/00), Influenza B virus (B/Nanchang/1/2000), Influenza B virus (B/Nanchang/12/98), Influenza B virus (B/Nanchang/15/95), Influenza B virus (B/Nanchang/15/97), Influenza B virus (B/Nanchang/195/94), Influenza B virus (B/Nanchang/

20

2/97), Influenza B virus (B/Nanchang/20/96), Influenza B virus (B/Nanchang/26/93), Influenza B virus (B/Nanchang/3/95), Influenza B virus (B/Nanchang/4/97), Influenza B virus (B/Nanchang/480/94), Influenza B virus (B/Nanchang/5/97), Influenza B virus (B/Nanchang/560/94), Influenza B virus (B/Nanchang/560a/94), Influenza B virus (B/Nanchang/560b/94), Influenza B virus (B/Nanchang/6/96), Influenza B virus (B/Nanchang/6/98), Influenza B virus (B/Nanchang/630/94), Influenza B virus (B/Nanchang/7/98), Influenza B virus (B/Nanchang/8/95), Influenza B virus (B/Nashville/107/93), Influenza B virus (B/Nashville/3/96), Influenza B virus (B/Nashville/34/96), Influenza B virus (B/Nashville/45/91), Influenza B virus (B/Nashville/48/91), Influenza B virus (B/Nashville/6/89), Influenza B virus (B/Nebraska/1/01), Influenza B virus (B/Nebraska/1/2005), Influenza B virus (B/Nebraska/2/01), Influenza B virus (B/Nebraska/4/2001), Influenza B virus (B/Nebraska/5/2003), Influenza B virus (B/Nepal/1078/2005), Influenza B virus (B/Nepal/1079/2005), Influenza B virus (B/Nepal/1080/2005), Influenza B virus (B/Nepal/1087/2005), Influenza B virus (B/Nepal/1088/2005), Influenza B virus (B/Nepal/1090/2005), Influenza B virus (B/Nepal/1092/2005), Influenza B virus (B/Nepal/1098/2005), Influenza B virus (B/Nepal/1101/2005), Influenza B virus (B/Nepal/1103/2005), Influenza B virus (B/Nepal/1104/2005), Influenza B virus (B/Nepal/1105/2005), Influenza B virus (B/Nepal/1106/2005), Influenza B virus (B/Nepal/1108/2005), Influenza B virus (B/Nepal/1114/2005), Influenza B virus (B/Nepal/1117/2005), Influenza B virus (B/Nepal/1118/2005), Influenza B virus (B/Nepal/1120/2005), Influenza B virus (B/Nepal/1122/2005), Influenza B virus (B/Nepal/1131/2005), Influenza B virus (B/Nepal/1132/2005), Influenza B virus (B/Nepal/1136/2005), Influenza B virus (B/Nepal/1137/2005), Influenza B virus (B/Nepal/1138/2005), Influenza B virus (B/Nepal/1139/2005), Influenza B virus (B/Nepal/1331/2005), Influenza B virus (B/Netherland/2781/90), Influenza B virus (B/Netherland/6357/90), Influenza B virus (B/Netherland/800/90), Influenza B virus (B/Netherland/801/90), Influenza B virus (B/Netherlands/1/97), Influenza B virus (B/Netherlands/13/94), Influenza B virus (B/Netherlands/2/95), Influenza B virus (B/Netherlands/31/95), Influenza B virus (B/Netherlands/32/94), Influenza B virus (B/Netherlands/384/95), Influenza B virus (B/Netherlands/429/98), Influenza B virus (B/Netherlands/580/89), Influenza B virus (B/Netherlands/6/96), Influenza B virus (B/Nevada/1/2001), Influenza B virus (B/Nevada/1/2002), Influenza B virus (B/Nevada/1/2005), Influenza B virus (B/Nevada/2/2003), Influenza B virus (B/Nevada/2/2006), Influenza B virus (B/Nevada/3/2006), Influenza B virus (B/Nevada/5/2005), Influenza B virus (B/New Jersey/1/2002), Influenza B virus (B/New Jersey/1/2004), Influenza B virus (B/New Jersey/1/2005), Influenza B virus (B/New Jersey/1/2006), Influenza B virus (B/New Jersey/3/2001), Influenza B virus (B/New Jersey/3/2005), Influenza B virus (B/New Jersey/4/2001), Influenza B virus (B/New Jersey/5/2005), Influenza B virus (B/New Jersey/6/2005), Influenza B virus (B/New Mexico/1/2001), Influenza B virus (B/New Mexico/1/2006), Influenza B virus (B/New Mexico/2/2005), Influenza B virus (B/New Mexico/9/2003), Influenza B virus (B/New York/1/2001), Influenza B virus (B/New York/1/2002), Influenza B virus (B/New York/1/2004), Influenza B virus (B/New York/1/2006), Influenza B virus (B/New York/10/2002), Influenza B virus (B/New York/11/2005), Influenza B virus (B/New York/12/2001), Influenza B virus (B/New York/12/2005), Influenza B virus (B/New York/12e/2005), Influenza B virus (B/New

York/14e/2005), Influenza B virus (B/New York/17/2004), Influenza B virus (B/New York/18/2003), Influenza B virus (B/New York/19/2004), Influenza B virus (B/New York/2/2000), Influenza B virus (B/New York/2/2002), Influenza B virus (B/New York/2/2006), Influenza B virus (B/New York/20139/99), Influenza B virus (B/New York/24/1993), Influenza B virus (B/New York/2e/2005), Influenza B virus (B/New York/3/90), Influenza B virus (B/New York/39/1991), Influenza B virus (B/New York/40/2002), Influenza B virus (B/New York/47/2001), Influenza B virus (B/New York/6/2004), Influenza B virus (B/New York/7/2002), Influenza B virus (B/New York/8/2000), Influenza B virus (B/New York/9/2002), Influenza B virus (B/New York/9/2004), Influenza B virus (B/New York/C10/2004), Influenza B virus (B/NIB/48/90), Influenza B virus (B/Ningxia/45/83), Influenza B virus (B/North Carolina/1/2005), Influenza B virus (B/North Carolina/3/2005), Influenza B virus (B/North Carolina/4/2004), Influenza B virus (B/North Carolina/5/2004), Influenza B virus (B/Norway/1/84), Influenza B virus (B/Ohio/1/2005), Influenza B virus (B/Ohio/1/X-19/2005), Influenza B virus (B/Ohio/1e/2005), Influenza B virus (B/Ohio/1e4/2005), Influenza B virus (B/Ohio/2/2002), Influenza B virus (B/Ohio/2e/2005), Influenza B virus (B/Oita/15/1992), Influenza B virus (B/Oklahoma/1/2006), Influenza B virus (B/Oklahoma/2/2005), Influenza B virus (B/Oman/16291/2001), Influenza B virus (B/Oman/16296/2001), Influenza B virus (B/Oman/16305/2001), Influenza B virus (B/Oregon/1/2005), Influenza B virus (B/Oregon/1/2006), Influenza B virus (B/Oregon/5/80), Influenza B virus (B/Osaka/1036/97), Influenza B virus (B/Osaka/1058/97), Influenza B virus (B/Osaka/1059/97), Influenza B virus (B/Osaka/1146/1997), Influenza B virus (B/Osaka/1169/97), Influenza B virus (B/Osaka/1201/2000), Influenza B virus (B/Osaka/547/1997), Influenza B virus (B/Osaka/547/97), Influenza B virus (B/Osaka/710/1997), Influenza B virus (B/Osaka/711/97), Influenza B virus (B/Osaka/728/1997), Influenza B virus (B/Osaka/755/1997), Influenza B virus (B/Osaka/820/1997), Influenza B virus (B/Osaka/837/1997), Influenza B virus (B/Osaka/854/1997), Influenza B virus (B/Osaka/983/1997), Influenza B virus (B/Osaka/983/1997-M1), Influenza B virus (B/Osaka/983/1997-M2), Influenza B virus (B/Osaka/983/97-V1), Influenza B virus (B/Osaka/983/97-V2), Influenza B virus (B/Osaka/983/97-V3), Influenza B virus (B/Osaka/983/97-V4), Influenza B virus (B/Osaka/983/97-V5), Influenza B virus (B/Osaka/983/97-V6), Influenza B virus (B/Osaka/983/97-V7), Influenza B virus (B/Osaka/c19/93), Influenza B virus (B/Oslo/1072/2001), Influenza B virus (B/Oslo/1329/2002), Influenza B virus (B/Oslo/1510/2002), Influenza B virus (B/Oslo/1846/2002), Influenza B virus (B/Oslo/1847/2002), Influenza B virus (B/Oslo/1862/2001), Influenza B virus (B/Oslo/1864/2001), Influenza B virus (B/Oslo/1870/2002), Influenza B virus (B/Oslo/1871/2002), Influenza B virus (B/Oslo/2293/2001), Influenza B virus (B/Oslo/2295/2001), Influenza B virus (B/Oslo/2297/2001), Influenza B virus (B/Oslo/238/2001), Influenza B virus (B/Oslo/3761/2000), Influenza B virus (B/Oslo/47/2001), Influenza B virus (B/Oslo/668/2002), Influenza B virus (B/Oslo/71/04), Influenza B virus (B/Oslo/801/99), Influenza B virus (B/Oslo/805/99), Influenza B virus (B/Oslo/837/99), Influenza B virus (B/Panama/45/1990), Influenza B virus (B/Panama/45/90), Influenza B virus (B/Paraguay/636/2003), Influenza B virus (B/Paris/329/90), Influenza B virus (B/Paris/549/1999), Influenza B virus (B/Parma/1/03), Influenza B virus (B/Parma/1/04), Influenza B virus (B/Parma/13/02), Influenza B virus

(B/Parma/16/02), Influenza B virus (B/Parma/2/03), Influenza B virus (B/Parma/2/04), Influenza B virus (B/Parma/23/02), Influenza B virus (B/Parma/24/02), Influenza B virus (B/Parma/25/02), Influenza B virus (B/Parma/28/02), Influenza B virus (B/Parma/3/04), Influenza B virus (B/Parma/4/04), Influenza B virus (B/Parma/5/02), Influenza B virus (B/Pennsylvania/1/2006), Influenza B virus (B/Pennsylvania/2/2001), Influenza B virus (B/Pennsylvania/2/2006), Influenza B virus (B/Pennsylvania/3/2003), Influenza B virus (B/Pennsylvania/3/2006), Influenza B virus (B/Pennsylvania/4/2004), Influenza B virus (B/Perth/211/2001), Influenza B virus (B/Perth/25/2002), Influenza B virus (B/Peru/1324/2004), Influenza B virus (B/Peru/1364/2004), Influenza B virus (B/Perugia/4/03), Influenza B virus (B/Philippines/5072/2001), Influenza B virus (B/Philippines/93079/2001), Influenza B virus (B/Pusan/250/99), Influenza B virus (B/Pusan/255/99), Influenza B virus (B/Pusan/270/99), Influenza B virus (B/Pusan/285/99), Influenza B virus (B/Quebec/1/01), Influenza B virus (B/Quebec/162/98), Influenza B virus (B/Quebec/173/98), Influenza B virus (B/Quebec/2/01), Influenza B virus (B/Quebec/3/01), Influenza B virus (B/Quebec/4/01), Influenza B virus (B/Quebec/452/98), Influenza B virus (B/Quebec/453/98), Influenza B virus (B/Quebec/465/98), Influenza B virus (B/Quebec/51/98), Influenza B virus (B/Quebec/511/98), Influenza B virus (B/Quebec/517/98), Influenza B virus (B/Quebec/6/01), Influenza B virus (B/Quebec/7/01), Influenza B virus (B/Quebec/74199/99), Influenza B virus (B/Quebec/74204/99), Influenza B virus (B/Quebec/74206/99), Influenza B virus (B/Quebec/8/01), Influenza B virus (B/Quebec/9/01), Influenza B virus (B/Rabat/41/97), Influenza B virus (B/Rabat/45/97), Influenza B virus (B/Rabat/61/97), Influenza B virus (B/RioJaneiro/200/02), Influenza B virus (B/RioJaneiro/209/02), Influenza B virus (B/RioJaneiro/315/01), Influenza B virus (B/RioJaneiro/353/02), Influenza B virus (B/RioGdoSul/337/01), Influenza B virus (B/RioGdoSul/357/02), Influenza B virus (B/RioGdoSul/374/01), Influenza B virus (B/Roma/1/03), Influenza B virus (B/Roma/2/03), Influenza B virus (B/Roma/3/03), Influenza B virus (B/Roma/4/02), Influenza B virus (B/Roma/7/02), Influenza B virus (B/Romania/217/1999), Influenza B virus (B/Romania/318/1998), Influenza B virus (B/Russia/22/1995), Influenza B virus (B/Saga/S172/99), Influenza B virus (B/Seal/Netherlands/1/99), Influenza B virus (B/Seoul/1/89), Influenza B virus (B/Seoul/1163/2004), Influenza B virus (B/Seoul/12/88), Influenza B virus (B/seoul/12/95), Influenza B virus (B/Seoul/13/95), Influenza B virus (B/Seoul/16/97), Influenza B virus (B/Seoul/17/95), Influenza B virus (B/Seoul/19/97), Influenza B virus (B/Seoul/21/95), Influenza B virus (B/Seoul/232/2004), Influenza B virus (B/Seoul/28/97), Influenza B virus (B/Seoul/31/97), Influenza B virus (B/Seoul/37/91), Influenza B virus (B/Seoul/38/91), Influenza B virus (B/Seoul/40/91), Influenza B virus (B/Seoul/41/91), Influenza B virus (B/Seoul/6/88), Influenza B virus (B/Shandong/7/97), Influenza B virus (B/Shangdong/7/97), Influenza B virus (B/Shanghai/1/77), Influenza B virus (B/Shanghai/10/80), Influenza B virus (B/Shanghai/24/76), Influenza B virus (B/Shanghai/35/84), Influenza B virus (B/Shanghai/361/03), Influenza B virus (B/Shanghai/361/2002), Influenza B virus (B/Shen-zhen/423/99), Influenza B virus (B/Shiga/51/98), Influenza B virus (B/Shiga/N18/98), Influenza B virus (B/Shiga/T30/98), Influenza B virus (B/Shiga/T37/98), Influenza B virus (B/Shizuoka/15/2001), Influenza B virus (B/Shizuoka/480/2000), Influenza B virus (B/Sichuan/281/96), Influenza B

virus (B/Sichuan/317/2001), Influenza B virus (B/Sichuan/379/99), Influenza B virus (B/Sichuan/38/2000), Influenza B virus (B/Sichuan/8/92), Influenza B virus (B/Siena/1/02), Influenza B virus (B/Singapore/04/1991), Influenza B virus (B/Singapore/11/1994), Influenza B virus (B/Singapore/22/1998), Influenza B virus (B/Singapore/222/79), Influenza B virus (B/Singapore/31/1998), Influenza B virus (B/Singapore/35/1998), Influenza B virus (B/South Australia/5/1999), Influenza B virus (B/South Carolina/04/2003), Influenza B virus (B/South Carolina/25723/99), Influenza B virus (B/South Carolina/3/2003), Influenza B virus (B/South Carolina/4/2003), Influenza B virus (B/South Dakota/1/2000), Influenza B virus (B/South Dakota/3/2003), Influenza B virus (B/South Dakota/5/89), Influenza B virus (B/Spain/WV22/2002), Influenza B virus (B/Spain/WV26/2002), Influenza B virus (B/Spain/WV27/2002), Influenza B virus (B/Spain/WV29/2002), Influenza B virus (B/Spain/WV33/2002), Influenza B virus (B/Spain/WV34/2002), Influenza B virus (B/Spain/WV36/2002), Influenza B virus (B/Spain/WV41/2002), Influenza B virus (B/Spain/WV42/2002), Influenza B virus (B/Spain/WV43/2002), Influenza B virus (B/Spain/WV45/2002), Influenza B virus (B/Spain/WV50/2002), Influenza B virus (B/Spain/WV51/2002), Influenza B virus (B/Spain/WV56/2002), Influenza B virus (B/Spain/WV57/2002), Influenza B virus (B/Spain/WV66/2002), Influenza B virus (B/Spain/WV69/2002), Influenza B virus (B/Spain/WV70/2002), Influenza B virus (B/Spain/WV73/2002), Influenza B virus (B/Spain/WV78/2002), Influenza B virus (B/St. Petersburg/14/2006), Influenza B virus (B/StaCatarina/308/02), Influenza B virus (B/StaCatarina/315/02), Influenza B virus (B/StaCatarina/318/02), Influenza B virus (B/StaCatarina/345/02), Influenza B virus (B/Stockholm/10/90), Influenza B virus (B/Suzuka/18/2005), Influenza B virus (B/Suzuka/28/2005), Influenza B virus (B/Suzuka/32/2005), Influenza B virus (B/Suzuka/58/2005), Influenza B virus (B/Switzerland/4291/97), Influenza B virus (B/Switzerland/5219/90), Influenza B virus (B/Switzerland/5241/90), Influenza B virus (B/Switzerland/5441/90), Influenza B virus (B/Switzerland/5444/90), Influenza B virus (B/Switzerland/5812/90), Influenza B virus (B/Switzerland/6121/90), Influenza B virus (B/Taiwan/0002/03), Influenza B virus (B/Taiwan/0114/01), Influenza B virus (B/Taiwan/0202/01), Influenza B virus (B/Taiwan/0409/00), Influenza B virus (B/Taiwan/0409/02), Influenza B virus (B/Taiwan/0562/03), Influenza B virus (B/Taiwan/0569/03), Influenza B virus (B/Taiwan/0576/03), Influenza B virus (B/Taiwan/0600/02), Influenza B virus (B/Taiwan/0610/03), Influenza B virus (B/Taiwan/0615/03), Influenza B virus (B/Taiwan/0616/03), Influenza B virus (B/Taiwan/0654/02), Influenza B virus (B/Taiwan/0684/03), Influenza B virus (B/Taiwan/0699/03), Influenza B virus (B/Taiwan/0702/02), Influenza B virus (B/Taiwan/0722/02), Influenza B virus (B/Taiwan/0730/02), Influenza B virus (B/Taiwan/0735/03), Influenza B virus (B/Taiwan/0833/03), Influenza B virus (B/Taiwan/0874/02), Influenza B virus (B/Taiwan/0879/02), Influenza B virus (B/Taiwan/0880/02), Influenza B virus (B/Taiwan/0927/02), Influenza B virus (B/Taiwan/0932/02), Influenza B virus (B/Taiwan/0993/02), Influenza B virus (B/Taiwan/1013/02), Influenza B virus (B/Taiwan/1013/03), Influenza B virus (B/Taiwan/102/2005), Influenza B virus (B/Taiwan/103/2005), Influenza B virus (B/Taiwan/110/2005), Influenza B virus (B/Taiwan/1103/2001), Influenza B virus (B/Taiwan/114/2001), Influenza B virus (B/Taiwan/1151/2001), Influenza B virus (B/Taiwan/117/2005), Influenza B virus (B/Taiwan/1197/1994), Influenza B virus (B/Taiwan/121/2005), Influenza B

virus (B/Taiwan/12192/2000), Influenza B virus (B/Taiwan/1243/99), Influenza B virus (B/Taiwan/1265/2000), Influenza B virus (B/Taiwan/1293/2000), Influenza B virus (B/Taiwan/13/2004), Influenza B virus (B/Taiwan/14/2004), 5 Influenza B virus (B/Taiwan/1484/2001), Influenza B virus (B/Taiwan/1502/02), Influenza B virus (B/Taiwan/1503/02), Influenza B virus (B/Taiwan/1534/02), Influenza B virus (B/Taiwan/1536/02), Influenza B virus (B/Taiwan/1561/02), Influenza B virus (B/Taiwan/1574/03), Influenza B virus 10 (B/Taiwan/1584/02), Influenza B virus (B/Taiwan/16/2004), Influenza B virus (B/Taiwan/1618/03), Influenza B virus (B/Taiwan/165/2005), Influenza B virus (B/Taiwan/166/2005), Influenza B virus (B/Taiwan/188/2005), Influenza B virus (B/Taiwan/1949/02), Influenza B virus (B/Taiwan/1950/02), Influenza B virus (B/Taiwan/202/2001), Influenza B virus (B/Taiwan/2027/99), Influenza B virus (B/Taiwan/217/97), Influenza B virus (B/Taiwan/21706/97), Influenza B virus (B/Taiwan/2195/99), Influenza B virus (B/Taiwan/2551/03), Influenza 15 B virus (B/Taiwan/2805/01), Influenza B virus (B/Taiwan/2805/2001), Influenza B virus (B/Taiwan/3143/97), Influenza B virus (B/Taiwan/31511/00), Influenza B virus (B/Taiwan/31511/2000), Influenza B virus (B/Taiwan/34/2004), Influenza B virus (B/Taiwan/3532/03), Influenza B virus 20 (B/Taiwan/39/2004), Influenza B virus (B/Taiwan/4101/00), Influenza B virus (B/Taiwan/4101/2000), Influenza B virus (B/Taiwan/4119/02), Influenza B virus (B/Taiwan/4184/00), Influenza B virus (B/Taiwan/4184/2000), Influenza B virus (B/Taiwan/43/2005), Influenza B virus (B/Taiwan/4602/02), Influenza B virus (B/Taiwan/473/2005), Influenza B virus (B/Taiwan/52/2004), Influenza B virus (B/Taiwan/52/2005), Influenza B virus (B/Taiwan/54/2004), Influenza B virus (B/Taiwan/61/2004), Influenza B virus 25 (B/Taiwan/635/2005), Influenza B virus (B/Taiwan/637/2005), Influenza B virus (B/Taiwan/68/2004), Influenza B virus (B/Taiwan/68/2005), Influenza B virus (B/Taiwan/69/2004), Influenza B virus (B/Taiwan/70/2005), Influenza B virus (B/Taiwan/74/2004), Influenza B virus (B/Taiwan/75/2004), Influenza B virus (B/Taiwan/77/2005), Influenza B 30 virus (B/Taiwan/81/2005), Influenza B virus (B/Taiwan/872/2005), Influenza B virus (B/Taiwan/97271/2001), Influenza B virus (B/Taiwan/98/2005), Influenza B virus (B/Taiwan/H96/02), Influenza B virus (B/Taiwan/M214/05), Influenza B virus (B/Taiwan/M227/05), Influenza B virus (B/Taiwan/45/04), Influenza B virus (B/Taiwan/M244/05), Influenza B virus (B/Taiwan/M251/05), Influenza B virus (B/Taiwan/M53/05), Influenza B virus (B/Taiwan/M71/01), Influenza B 35 virus (B/Taiwan/N1013/99), Influenza B virus (B/Taiwan/N1115/02), Influenza B virus (B/Taiwan/N1207/99), Influenza B virus (B/Taiwan/N1316/01), Influenza B virus (B/Taiwan/N1549/01), Influenza B virus (B/Taiwan/N1582/02), Influenza B virus (B/Taiwan/N16/03), Influenza B virus 40 (B/Taiwan/N1619/04), Influenza B virus (B/Taiwan/N1848/02), Influenza B virus (B/Taiwan/N1902/04), Influenza B virus (B/Taiwan/N200/05), Influenza B virus (B/Taiwan/N2050/02), Influenza B virus (B/Taiwan/N230/01), Influenza B virus (B/Taiwan/N232/00), Influenza B virus (B/Taiwan/N233/02), Influenza B virus (B/Taiwan/N2335/01), Influenza B virus (B/Taiwan/N253/03), Influenza B virus 45 (B/Taiwan/N260/04), Influenza B virus (B/Taiwan/N2986/02), Influenza B virus (B/Taiwan/N3688/04), Influenza B virus (B/Taiwan/N371/05), Influenza B virus (B/Taiwan/N376/05), Influenza B virus (B/Taiwan/N384/03), Influenza B virus (B/Taiwan/N3849/02), Influenza B virus (B/Taiwan/N404/02), Influenza B virus (B/Taiwan/N473/00), Influenza B virus (B/Taiwan/N511/01), Influenza B virus (B/Taiwan/N559/05), Influenza B virus (B/Taiwan/N612/01), Influenza B 50

25

B virus (B/Taiwan/N701/01), Influenza B virus (B/Taiwan/N767/01), Influenza B virus (B/Taiwan/N798/05), Influenza B virus (B/Taiwan/N860/05), Influenza B virus (B/Taiwan/N872/04), Influenza B virus (B/Taiwan/N913/04), Influenza B virus (B/Taiwan/S117/05), Influenza B virus (B/Taiwan/S141/02), Influenza B virus (B/Taiwan/S76/02), Influenza B virus (B/Taiwan/S82/02), Influenza B virus (B/Taiwan/103/2005), Influenza B virus (B/Tehran/80/02), Influenza B virus (B/Temple/B10/1999), Influenza B virus (B/Temple/B1166/2001), Influenza B virus (B/Temple/B1181/2001), Influenza B virus (B/Temple/B1182/2001), Influenza B virus (B/Temple/B1188/2001), Influenza B virus (B/Temple/B1190/2001), Influenza B virus (B/Temple/B1193/2001), Influenza B virus (B/Temple/B17/2003), Influenza B virus (B/Temple/B18/2003), Influenza B virus (B/Temple/B19/2003), Influenza B virus (B/Temple/B20/2003), Influenza B virus (B/Temple/B21/2003), Influenza B virus (B/Temple/B24/2003), Influenza B virus (B/Temple/B3/1999), Influenza B virus (B/Temple/B30/2003), Influenza B virus (B/Temple/B7/1999), Influenza B virus (B/Temple/B8/1999), Influenza B virus (B/Temple/B9/1999), Influenza B virus (B/Texas/06/2000), Influenza B virus (B/Texas/1/2000), Influenza B virus (B/Texas/1/2004), Influenza B virus (B/Texas/1/2006), Influenza B virus (B/Texas/1/91), Influenza B virus (B/Texas/10/2005), Influenza B virus (B/Texas/11/2001), Influenza B virus (B/Texas/12/2001), Influenza B virus (B/Texas/14/2001), Influenza B virus (B/Texas/16/2001), Influenza B virus (B/Texas/18/2001), Influenza B virus (B/Texas/2/2006), Influenza B virus (B/Texas/22/2001), Influenza B virus (B/Texas/23/2000), Influenza B virus (B/Texas/3/2001), Influenza B virus (B/Texas/3/2002), Influenza B virus (B/Texas/3/2006), Influenza B virus (B/Texas/37/1988), Influenza B virus (B/Texas/37/88), Influenza B virus (B/Texas/4/2006), Influenza B virus (B/Texas/4/90), Influenza B virus (B/Texas/5/2002), Influenza B virus (B/Texas/57/2002), Influenza B virus (B/Texas/6/2000), Influenza B virus (B/Tokushima/101/93), Influenza B virus (B/Tokyo/6/98), Influenza B virus (B/Trento/3/02), Influenza B virus (B/Trieste/1/02), Influenza B virus (B/Trieste/1/03), Influenza B virus (B/Trieste/17/02), Influenza B virus (B/Trieste/19/02), Influenza B virus (B/Trieste/2/03), Influenza B virus (B/Trieste/25/02), Influenza B virus (B/Trieste/27/02), Influenza B virus (B/Trieste/28/02), Influenza B virus (B/Trieste/34/02), Influenza B virus (B/Trieste/37/02), Influenza B virus (B/Trieste/4/02), Influenza B virus (B/Trieste/8/02), Influenza B virus (B/Trieste/14/02), Influenza B virus (B/Trieste/18/02), Influenza B virus (B/Trieste/23/02), Influenza B virus (B/Trieste/7/02), Influenza B virus (B/Ulan Ude/4/02), Influenza B virus (B/Ulan-Ude/6/2003), Influenza B virus (B/UlanUde/4/02), Influenza B virus (B/United Kingdom/34304/99), Influenza B virus (B/United Kingdom/34520/99), Influenza B virus (B/Uruguay/19/02), Influenza B virus (B/Uruguay/19/05), Influenza B virus (B/Uruguay/2/02), Influenza B virus (B/Uruguay/28/05), Influenza B virus (B/Uruguay/33/05), Influenza B virus (B/Uruguay/4/02), Influenza B virus (B/Uruguay/5/02), Influenza B virus (B/Uruguay/65/05), Influenza B virus (B/Uruguay/7/02), Influenza B virus (B/Uruguay/74/04), Influenza B virus (B/Uruguay/75/04), Influenza B virus (B/Uruguay/NG/02), Influenza B virus (B/Ushuaia/15732/99), Influenza B virus (B/USSR/100/83), Influenza B virus (B/Utah/1/2005), Influenza B virus (B/Utah/20139/99), Influenza B virus (B/Utah/20975/99), Influenza B virus (B/Vermont/1/2006), Influenza B virus (B/Victoria/02/1987), Influenza B virus (B/Victoria/103/89),

26

Influenza B virus (B/Victoria/19/89), Influenza B virus (B/Victoria/2/87), Influenza B virus (B/Victoria/504/2000), Influenza B virus (B/Vienna/1/99), Influenza B virus (B/Virginia/1/2005), Influenza B virus (B/Virginia/1/2006), Influenza B virus (B/Virginia/11/2003), Influenza B virus (B/Virginia/2/2006), Influenza B virus (B/Virginia/3/2003), Influenza B virus (B/Virginia/3/2006), Influenza B virus (B/Virginia/9/2005), Influenza B virus (B/Washington/)/2004), Influenza B virus (B/Washington/2/2000), Influenza B virus (B/Washington/2/2004), Influenza B virus (B/Washington/3/2000), Influenza B virus (B/Washington/3/2003), Influenza B virus (B/Washington/5/2005), Influenza B virus (B/Wellington/01/1994), Influenza B virus (B/Wisconsin/1/2004), Influenza B virus (B/Wisconsin/1/2006), Influenza B virus (B/Wisconsin/10/2006), Influenza B virus (B/Wisconsin/15e/2005), Influenza B virus (B/Wisconsin/17/2006), Influenza B virus (B/Wisconsin/2/2004), Influenza B virus (B/Wisconsin/2/2006), Influenza B virus (B/Wisconsin/22/2006), Influenza B virus (B/Wisconsin/26/2006), Influenza B virus (B/Wisconsin/29/2006), Influenza B virus (B/Wisconsin/3/2000), Influenza B virus (B/Wisconsin/3/2004), Influenza B virus (B/Wisconsin/3/2006), Influenza B virus (B/Wisconsin/31/2006), Influenza B virus (B/Wisconsin/4/2006), Influenza B virus (B/Wisconsin/5/2006), Influenza B virus (B/Wisconsin/6/2006), Influenza B virus (B/Wisconsin/7/2002), Influenza B virus (B/Wuhan/2/2001), Influenza B virus (B/Wuhan/356/2000), Influenza B virus (B/WV194/2002), Influenza B virus (B/Wyoming/16/2001), Influenza B virus (B/Wyoming/2/2003), Influenza B virus (B/Xuanwu/1/82), Influenza B virus (B/Xuanwu/23/82), Influenza B virus (B/Yamagata/1/73), Influenza B virus (B/Yamagata/115/2003), Influenza B virus (B/Yamagata/1246/2003), Influenza B virus (B/Yamagata/1311/2003), Influenza B virus (B/Yamagata/16/1988), Influenza B virus (B/Yamagata/16/88), Influenza B virus (B/Yamagata/222/2002), Influenza B virus (B/Yamagata/K198/2001), Influenza B virus (B/Yamagata/K246/2001), Influenza B virus (B/Yamagata/K270/2001), Influenza B virus (B/Yamagata/K298/2001), Influenza B virus (B/Yamagata/K320/2001), Influenza B virus (B/Yamagata/K354/2001), Influenza B virus (B/Yamagata/K386/2001), Influenza B virus (B/Yamagata/K411/2001), Influenza B virus (B/Yamagata/K461/2001), Influenza B virus (B/Yamagata/K490/2001), Influenza B virus (B/Yamagata/K500/2001), Influenza B virus (B/Yamagata/K501/2001), Influenza B virus (B/Yamagata/K508/2001), Influenza B virus (B/Yamagata/K513/2001), Influenza B virus (B/Yamagata/K515/2001), Influenza B virus (B/Yamagata/K519/2001), Influenza B virus (B/Yamagata/K520/2001), Influenza B virus (B/Yamagata/K521/2001), Influenza B virus (B/Yamagata/K535/2001), Influenza B virus (B/Yamagata/K542/2001), Influenza B virus (B/Yamanashi/166/1998), Influenza B virus (B/Yamanashi/166/98), Influenza B virus (B/Yunnan/123/2001), Influenza B virus (strain B/Alaska/12/96), Influenza B virus (STRAIN B/ANN ARBOR/1/66 [COLD-ADAPTED]), Influenza B virus (STRAIN B/ANN ARBOR/1/66 [WILD-TYPE]), Influenza B virus (STRAIN B/BA/78), Influenza B virus (STRAIN B/BEIJING/1/87), Influenza B virus (STRAIN B/ENGLAND/222/82), Influenza B virus (strain B/finland/145/90), Influenza B virus (strain B/finland/146/90), Influenza B virus (strain B/finland/147/90), Influenza B virus (strain B/finland/148/90), Influenza B virus (strain B/finland/149/90), Influenza B virus (strain B/finland/150/90), Influenza B virus (strain B/finland/151/90), Influenza B virus (strain B/finland/24/85), Influenza B virus (strain

B/finland/56/88), Influenza B virus (STRAIN B/FU-KUOKA/80/81), Influenza B virus (STRAIN B/GA/86), Influenza B virus (STRAIN B/GL/54), Influenza B virus (STRAIN B/HONG KONG/8/73), Influenza B virus (STRAIN B/HT/84), Influenza B virus (STRAIN B/ID/86), Influenza B virus (STRAIN B/LENINGRAD/179/86), Influenza B virus (STRAIN B/MARYLAND/59), Influenza B virus (STRAIN B/MEMPHIS/6/86), Influenza B virus (STRAIN B/NAGASAKI/1/87), Influenza B virus (strain B/Osaka/491/97), Influenza B virus (STRAIN B/PA/79), Influenza B virus (STRAIN B/RU/69), Influenza B virus (STRAIN B/SINGAPORE/64), Influenza B virus (strain B/Tokyo/942/96), Influenza B virus (STRAIN B/VICTORIA/3/85), Influenza B virus (STRAIN B/VICTORIA/87), Influenza B virus(B/Rochester/02/2001), and other subtypes. In further embodiments, the influenza virus C belongs to but is not limited to subtype Influenza C virus (C/Aichi/1/81), Influenza C virus (C/Aichi/1/99), Influenza C virus (C/Ann Arbor/1/50), Influenza C virus (C/Aomori/74), Influenza C virus (C/California/78), Influenza C virus (C/England/83), Influenza C virus (C/Fukuoka/2/2004), Influenza C virus (C/Fukuoka/3/2004), Influenza C virus (C/Fukushima/1/2004), Influenza C virus (C/Greece/79), Influenza C virus (C/Hiroshima/246/2000), Influenza C virus (C/Hiroshima/247/2000), Influenza C virus (C/Hiroshima/248/2000), Influenza C virus (C/Hiroshima/249/2000), Influenza C virus (C/Hiroshima/250/2000), Influenza C virus (C/Hiroshima/251/2000), Influenza C virus (C/Hiroshima/252/2000), Influenza C virus (C/Hiroshima/252/99), Influenza C virus (C/Hiroshima/290/99), Influenza C virus (C/Hiroshima/4/2004), Influenza C virus (C/Hyogo/1/83), Influenza C virus (C/Johannesburg/1/66), Influenza C virus (C/Johannesburg/66), Influenza C virus (C/Kanagawa/1/76), Influenza C virus (C/Kanagawa/2/2004), Influenza C virus (C/Kansas/1/79), Influenza C virus (C/Kyoto/1/79), Influenza C virus (C/Kyoto/41/82), Influenza C virus (C/Mississippi/80), Influenza C virus (C/Miyagi/1/90), Influenza C virus (C/Miyagi/1/93), Influenza C virus (C/Miyagi/1/94), Influenza C virus (C/Miyagi/1/97), Influenza C virus (C/Miyagi/1/99), Influenza C virus (C/Miyagi/12/2004), Influenza C virus (C/Miyagi/2/2000), Influenza C virus (C/Miyagi/2/92), Influenza C virus (C/Miyagi/2/93), Influenza C virus (C/Miyagi/2/94), Influenza C virus (C/Miyagi/2/96), Influenza C virus (C/Miyagi/2/98), Influenza C virus (C/Miyagi/3/2000), Influenza C virus (C/Miyagi/3/91), Influenza C virus (C/Miyagi/3/92), Influenza C virus (C/Miyagi/3/93), Influenza C virus (C/Miyagi/3/94), Influenza C virus (C/Miyagi/3/97), Influenza C virus (C/Miyagi/3/99), Influenza C virus (C/Miyagi/4/2000), Influenza C virus (C/Miyagi/4/93), Influenza C virus (C/Miyagi/4/96), Influenza C virus (C/Miyagi/4/97), Influenza C virus (C/Miyagi/4/98), Influenza C virus (C/Miyagi/42/2004), Influenza C virus (C/Miyagi/5/2000), Influenza C virus (C/Miyagi/5/91), Influenza C virus (C/Miyagi/5/93), Influenza C virus (C/Miyagi/6/93), Influenza C virus (C/Miyagi/6/96), Influenza C virus (C/Miyagi/7/91), Influenza C virus (C/Miyagi/7/93), Influenza C virus (C/Miyagi/7/96), Influenza C virus (C/Miyagi/77), Influenza C virus (C/Miyagi/8/96), Influenza C virus (C/Miyagi/9/91), Influenza C virus (C/Miyagi/9/96), Influenza C virus (C/Nara/1/85), Influenza C virus (C/Nara/2/85), Influenza C virus (C/Nara/82), Influenza C virus (C/New Jersey/76), Influenza C virus (C/Niigata/1/2004), Influenza C virus (C/Osaka/2/2004), Influenza C virus (C/pig/Beijing/115/81), Influenza C virus (C/Saitama/1/2000), Influenza C virus (C/Saitama/1/2004), Influenza C virus (C/Saitama/2/2000), Influenza C virus (C/Saitama/3/2000), Influenza C virus (C/Sapporo/71), Influenza C virus

(C/Shizuoka/79), Influenza C virus (C/Yamagata/1/86), Influenza C virus (C/Yamagata/1/88), Influenza C virus (C/Yamagata/10/89), Influenza C virus (C/Yamagata/13/98), Influenza C virus (C/Yamagata/15/2004), Influenza C virus (C/Yamagata/2/2000), Influenza C virus (C/Yamagata/2/98), Influenza C virus (C/Yamagata/2/99), Influenza C virus (C/Yamagata/20/2004), Influenza C virus (C/Yamagata/20/96), Influenza C virus (C/Yamagata/21/2004), Influenza C virus (C/Yamagata/26/81), Influenza C virus (C/Yamagata/27/2004), Influenza C virus (C/Yamagata/3/2000), Influenza C virus (C/Yamagata/3/2004), Influenza C virus (C/Yamagata/3/88), Influenza C virus (C/Yamagata/3/96), Influenza C virus (C/Yamagata/4/88), Influenza C virus (C/Yamagata/4/89), Influenza C virus (C/Yamagata/5/92), Influenza C virus (C/Yamagata/6/2000), Influenza C virus (C/Yamagata/6/98), Influenza C virus (C/Yamagata/64), Influenza C virus (C/Yamagata/7/88), Influenza C virus (C/Yamagata/8/2000), Influenza C virus (C/Yamagata/8/88), Influenza C virus (C/Yamagata/8/96), Influenza C virus (C/Yamagata/9/2000), Influenza C virus (C/Yamagata/9/88), Influenza C virus (C/Yamagata/9/96), Influenza C virus (STRAIN C/BERLIN/1/85), Influenza C virus (STRAIN C/ENGLAND/892/83), Influenza C virus (STRAIN C/GREAT LAKES/1167/54), Influenza C virus (STRAIN C/J/50), Influenza C virus (STRAIN C/PIG/BEIJING/10/81), Influenza C virus (STRAIN C/PIG/BEIJING/439/82), Influenza C virus (STRAIN C/TAYLOR/1233/47), Influenza C virus (STRAIN C/YAMAGATA/10/81), Isavirus or Infectious salmon anemia virus, Thogotovirus or Dhori virus, Batken virus, Dhori virus (STRAIN INDIAN/1313/61) or Thogoto virus, Thogoto virus (isolate SiAr 126) or unclassified Thogotovirus, Araguari virus, unclassified Orthomyxoviridae or Fowl plague virus or Swine influenza virus or unidentified influenza virus and other subtypes.

In various embodiments, the attenuated virus belongs to the delta virus family and all related genera.

In various embodiments, the attenuated virus belongs to the Adenoviridae virus family and all related genera, strains, types and isolates for example but not limited to human adenovirus A, B C.

In various embodiments, the attenuated virus belongs to the Herpesviridae virus family and all related genera, strains, types and isolates for example but not limited to herpes simplex virus.

In various embodiments, the attenuated virus belongs to the Reoviridae virus family and all related genera, strains, types and isolates.

In various embodiments, the attenuated virus belongs to the Papillomaviridae virus family and all related genera, strains, types and isolates.

In various embodiments, the attenuated virus belongs to the Poxviridae virus family and all related genera, strains, types and isolates.

In various embodiments, the attenuated virus belongs to the Retroviridae virus family and all related genera, strains, types and isolates. For example but not limited to Human Immunodeficiency Virus.

In various embodiments, the attenuated virus belongs to the Filoviridae virus family and all related genera, strains, types and isolates.

In various embodiments, the attenuated virus belongs to the Paramyxoviridae virus family and all related genera, strains, types and isolates.

In various embodiments, the attenuated virus belongs to the Orthomyxoviridae virus family and all related genera, strains, types and isolates.

29

In various embodiments, the attenuated virus belongs to the Picornaviridae virus family and all related genera, strains, types and isolates.

In various embodiments, the attenuated virus belongs to the Bunyaviridae virus family and all related genera, strains, types and isolates.

In various embodiments, the attenuated virus belongs to the Nidovirales virus family and all related genera, strains, types and isolates.

In various embodiments, the attenuated virus belongs to the Caliciviridae virus family and all related genera, strains, types and isolates.

In certain embodiments, the synonymous codon substitutions alter codon bias, codon pair bias, density of deoptimized codons and deoptimized codon pairs, RNA secondary structure, CpG dinucleotide content, C+G content, translation frameshift sites, translation pause sites, the presence or absence microRNA recognition sequences or any combination thereof, in the genome. The codon substitutions may be engineered in multiple locations distributed throughout the genome, or in the multiple locations restricted to a portion of the genome. In further embodiments, the portion of the genome is the capsid coding region.

In preferred embodiments of this invention, the virus retains the ability to induce a protective immune response in an animal host. In other preferred embodiments, the virulence of the virus does not revert to wild type.

Poliovirus, Rhinovirus, and Influenza Virus

Poliovirus, a member of the Picornavirus family, is a small non-enveloped virus with a single stranded (+) sense RNA genome of 7.5 kb in length (Kitamura et al., 1981). Upon cell entry, the genomic RNA serves as an mRNA encoding a single polyprotein that after a cascade of autocatalytic cleavage events gives rise to full complement of functional poliovirus proteins. The same genomic RNA serves as a template for the synthesis of (-) sense RNA, an intermediary for the synthesis of new (+) strands that either serve as mRNA, replication template or genomic RNA destined for encapsidation into progeny virions (Mueller et al., 2005). As described herein, the well established PV system was used to address general questions of optimizing design strategies for the production of attenuated synthetic viruses. PV provides one of the most important and best understood molecular models for developing anti-viral strategies. In particular, a reverse genetics system exists whereby viral nucleic acid can be synthesized *in vitro* by completely synthetic methods and then converted into infectious virions (see below). Furthermore, a convenient mouse model is available (CD155tg mice, which express the human receptor for polio) for testing attenuation of synthetic PV designs as previously described (Cello et al., 2002).

Rhinoviruses are also members of the Picornavirus family, and are related to PV. Human Rhinoviruses (HRV) are the usual causative agent of the common cold, and as such they are responsible for more episodes of illness than any other infectious agent (Hendley, 1999). In addition to the common cold, HRV is also involved in ear and sinus infections, asthmatic attacks, and other diseases. Similar to PV, HRV comprises a single-stranded positive sense RNA virus, whose genome encodes a self-processing polyprotein. The RNA is translated through an internal initiation mechanism using an Internal Ribosome Entry Site (IRES) to produce structural proteins that form the capsid, as well as non-structural proteins such as the two viral proteases, 2A and 3C, and the RNA-dependent polymerase (Jang et al., 1989; Pelletier et al., 1988). Also like PV, HRV has a non-enveloped icosahedral capsid, formed by 60 copies of

30

the four capsid proteins VP1-4 (Savolainen et al., 2003). The replication cycle of HRV is also identical to that of poliovirus. The close similarity to PV, combined with the significant, almost ubiquitous impact on human health, makes HRV an extremely attractive candidate for generating a novel attenuated virus useful for immunization.

Despite decades of research by pharmaceutical companies, no successful drug against HRV has been developed. This is partly due to the relatively low risk tolerance of federal regulators and the public for drugs that treat a mostly non-serious infection. That is, even minor side effects are unacceptable. Thus, in the absence of a drug, there is a clear desire for a safe and effective anti-rhinovirus vaccine. However, developing an anti-rhinovirus vaccine is extremely challenging, because there are over 100 serotypes of HRV, of which approximately 30 circulate widely and infect humans regularly. An effective vaccine must enable the immune system to recognize every single serotype in order to confer true immunity. The SAVE approach described herein offers a practical solution to the development of an effective rhinovirus vaccine. Based on the predictability of the SAVE design process, it would be inexpensive to design and synthesize 100 or more SAVE-attenuated rhinoviruses, which in combination would constitute a vaccine.

Influenza virus—Between 1990 and 1999, influenza viruses caused approximately 35,000 deaths each year in the U.S.A. (Thompson et al., 2003). Together with approximately 200,000 hospitalizations, the impact on the U.S. economy has been estimated to exceed \$23 billion annually (Cram et al., 2001). Globally, between 300,000 to 500,000 people die each year due to influenza virus infections (Kamps et al., 2006). Although the virus causes disease amongst all age groups, the rates of serious complications are highest in children and persons over 65 years of age. Influenza has the potential to mutate or recombine into extremely deadly forms, as happened during the great influenza epidemic of 1918, in which about 30 million people died. This was possibly the single most deadly one-year epidemic in human history.

Influenza viruses are divided into three types A, B, and C. Antigenicity is determined by two glycoproteins at the surface of the enveloped virion: hemagglutinin (HA) and neuraminidase (NA). Both glycoproteins continuously change their antigenicity to escape humoral immunity. Altering the glycoproteins allows virus strains to continue infecting vaccinated individuals, which is the reason for yearly vaccination of high-risk groups. In addition, human influenza viruses can replace the HA or NA glycoproteins with those of birds and pigs, a reassortment of gene segments, known as genetic shift, leading to new viruses (H1N1 to H2N2 or H3N2, etc.) (Steinhauer and Skehel, 2002). These novel viruses, to which the global population is immunologically naive, are the cause of pandemics that kill millions of people (Kilbourne, 2006; Russell and Webster, 2005). The history of influenza virus, together with the current threat of the highly pathogenic avian influenza virus, H5N1 (Stephenson and Democratis, 2006), underscores the need for preventing influenza virus disease.

Currently, two influenza vaccines are in use: a live, attenuated vaccine (cold adapted; "FluMist") and an inactivated virus. The application of the attenuated vaccine is restricted to healthy children, adolescents and adults (excluding pregnant females), ages 5-49. This age restriction leaves out precisely those who are at highest risks of influenza. Furthermore, the attenuated FluMist virus has the possibility of reversion, which is usual for a live virus. Production of the second, more commonly administered

31

inactivated influenza virus vaccine is complex. Further, this vaccine appears to be less effective than hoped for in preventing death in the elderly (>65-year-old) population (Simonson et al., 2005). These facts underscore the need for novel strategies to generate influenza virus vaccines.

Reverse Genetics of Picornaviruses

Reverse genetics generally refers to experimental approaches to discovering the function of a gene that proceeds in the opposite direction to the so-called forward genetic approaches of classical genetics. That is, whereas forward genetics approaches seek to determine the function of a gene by elucidating the genetic basis of a phenotypic trait, strategies based on reverse genetics begin with an isolated gene and seek to discover its function by investigating the possible phenotypes generated by expression of the wt or mutated gene. As used herein in the context of viral systems, “reverse genetics” systems refer to the availability of techniques that permit genetic manipulation of viral genomes made of RNA. Briefly, the viral genomes are isolated from virions or from infected cells, converted to DNA (“cDNA”) by the enzyme reverse transcriptase, possibly modified as desired, and reverted, usually via the RNA intermediate, back into infectious viral particles. This process in picornaviruses is extremely simple; in fact, the first reverse genetics system developed for any animal RNA virus was for PV (Racaniello and Baltimore, 1981). Viral reverse genetics systems are based on the historical finding that naked viral genomic RNA is infectious when transfected into a suitable mammalian cell (Alexander et al., 1958). The discovery of reverse transcriptase and the development of molecular cloning techniques in the 1970’s enabled scientists to generate and manipulate cDNA copies of RNA viral genomes. Most commonly, the entire cDNA copy of the genome is cloned immediately downstream of a phage T7 RNA polymerase promoter that allows the in vitro synthesis of genome RNA, which is then transfected into cells for generation of virus (van der Wert, et al., 1986). Alternatively, the same DNA plasmid may be transfected into cells expressing the T7 RNA polymerase in the cytoplasm. This system can be used for various viral pathogens including both PV and HRV.

Molecular Virology and Reverse Genetics of Influenza Virus

Influenza virus, like the picornaviruses, PV and HRV, is an RNA virus, but is otherwise unrelated to and quite different from PV. In contrast to the picornaviruses, influenza is a minus strand virus. Furthermore, influenza consists of eight separate gene segments ranging from 890 to 2341 nucleotides (Lamb and Krug, 2001). Partly because of the minus strand organization, and partly because of the eight separate gene segments, the reverse genetics system is more complex than for PV. Nevertheless, a reverse genetics system has been developed for influenza virus (Enami et al., 1990; Fodor et al., 1999; Garcia-Sastre and Palese, 1993; Hoffman et al., 2000; Luytjes et al., 1989; Neumann et al., 1999). Each of the eight gene segments is expressed from a separate plasmid. This reverse genetics system is extremely convenient for use in the SAVE strategy described herein, because the longest individual gene segment is less than 3 kb, and thus easy to synthesize and manipulate. Further, the different gene segments can be combined and recombined simply by mixing different plasmids. Thus, application of SAVE methods are possibly even more feasible for influenza virus than for PV.

A recent paradigm shift in viral reverse genetics occurred with the present inventors’ first chemical synthesis of an infectious virus genome by assembly from synthetic DNA

32

oligonucleotides (Cello et al., 2002). This achievement made it clear that most or all viruses for which a reverse genetics system is available can be synthesized solely from their genomic sequence information, and promises unprecedented flexibility in re-synthesizing and modifying these viruses to meet desired criteria.

De Novo Synthesis of Viral Genomes

Computer-based algorithms are used to design and synthesize viral genomes de novo. These synthesized genomes, exemplified by the synthesis of attenuated PV described herein, encode exactly the same proteins as wild type(wt) viruses, but by using alternative synonymous codons, various parameters, including codon bias, codon pair bias, RNA secondary structure, and/or dinucleotide content, are altered. The presented data show that these coding-independent changes produce highly attenuated viruses, often due to poor translation of proteins. By targeting an elementary function of all viruses, namely protein translation, a very general method has been developed for predictably, safely, quickly and cheaply producing attenuated viruses, which are useful for making vaccines. This method, dubbed “SAVE” (Synthetic Attenuated Virus Engineering), is applicable to a wide variety of viruses other than PV for which there is a medical need for new vaccines. These viruses include, but are not limited to rhinovirus, influenza virus, SARS and other coronaviruses, HIV, HCV, infectious bronchitis virus, ebolavirus, Marburg virus, dengue fever virus, *West Nile* disease virus, EBV, yellow fever virus, enteroviruses other than poliovirus, such as echoviruses, coxsackie viruses, and entrovirus71; hepatitis A virus, aphthoviruses, such as foot-and-mouth-disease virus, myxoviruses, such as influenza viruses, paramyxoviruses, such as measles virus, mumps virus, respiratory syncytia virus, flaviviruses such as dengue virus, yellow fever virus, St. Louis encephalitis virus and tick-born virus, alphaviruses, such as Western- and Eastern encephalitis virus, hepatitis B virus, and bovine diarrhea virus, and ebolavirus.

Both codon and codon-pair deoptimization in the PV capsid coding region are shown herein to dramatically reduce PV fitness. The present invention is not limited to any particular molecular mechanism underlying virus attenuation via substitution of synonymous codons. Nevertheless, experiments are ongoing to better understand the underlying molecular mechanisms of codon and codon pair deoptimization in producing attenuated viruses. In particular, evidence is provided in this application that indicates that codon deoptimization and codon pair deoptimization can result in inefficient translation. High throughput methods for the quick generation and screening of large numbers of viral constructs are also being developed.

Large-Scale DNA Assembly

In recent years, the plunging costs and increasing quality of oligonucleotide synthesis have made it practical to assemble large segments of DNA (at least up to about 10 kb) from synthetic oligonucleotides. Commercial vendors such as Blue Heron Biotechnology, Inc. (Bothwell, Wash.) (and also many others) currently synthesize, assemble, clone, sequence-verify, and deliver a large segment of synthetic DNA of known sequence for the relatively low price of about \$1.50 per base. Thus, purchase of synthesized viral genomes from commercial suppliers is a convenient and cost-effective option, and prices continue to decrease rapidly. Furthermore, new methods of synthesizing and assembling very large DNA molecules at extremely low costs are emerging (Tian et al., 2004). The Church lab has pioneered a method that uses parallel synthesis of thousands of oligonucleotides (for instance, on photo-programmable microfluid-

idics chips, or on microarrays available from Nimblegen Systems, Inc., Madison, Wis., or Agilent Technologies, Inc., Santa Clara, Calif.), followed by error reduction and assembly by overlap PCR. These methods have the potential to reduce the cost of synthetic large DNAs to less than 1 cent per base. The improved efficiency and accuracy, and rapidly declining cost, of large-scale DNA synthesis provides an impetus for the development and broad application of the SAVE strategy.

Alternative Encoding, Codon Bias, and Codon Pair Bias Alternative Encoding

A given peptide can be encoded by a large number of nucleic acid sequences. For example, even a typical short 10-mer oligopeptide can be encoded by approximately 4^{10} (about 10^6) different nucleic acids, and the proteins of PV can be encoded by about 10^{442} different nucleic acids. Natural selection has ultimately chosen one of these possible 10^{442} nucleic acids as the PV genome. Whereas the primary amino acid sequence is the most important level of information encoded by a given mRNA, there are additional kinds of information within different kinds of RNA sequences. These include RNA structural elements of distinct function (e.g., for PV, the cis-acting replication element, or CRE (Goodfellow et al., 2000; McKnight, 2003), translational kinetic signals (pause sites, frame shift sites, etc.), polyadenylation signals, splice signals, enzymatic functions (ribozyme) and, quite likely, other as yet unidentified information and signals).

Even with the caveat that signals such as the CRE must be preserved, 10^{442} possible encoding sequences provide tremendous flexibility to make drastic changes in the RNA sequence of polio while preserving the capacity to encode the same protein. Changes can be made in codon bias or codon pair bias, and nucleic acid signals and secondary structures in the RNA can be added or removed. Additional or novel proteins can even be simultaneously encoded in alternative frames (see, e.g., Wang et al., 2006).

Codon Bias

Whereas most amino acids can be encoded by several different codons, not all codons are used equally frequently: some codons are “rare” codons, whereas others are “frequent” codons. As used herein, a “rare” codon is one of at least two synonymous codons encoding a particular amino acid that is present in an mRNA at a significantly lower frequency than the most frequently used codon for that amino acid. Thus, the rare codon may be present at about a 2-fold lower frequency than the most frequently used codon. Preferably, the rare codon is present at least a 3-fold, more preferably at least a 5-fold, lower frequency than the most frequently used codon for the amino acid. Conversely, a “frequent” codon is one of at least two synonymous codons encoding a particular amino acid that is present in an mRNA at a significantly higher frequency than the least frequently used codon for that amino acid. The frequent codon may be present at about a 2-fold, preferably at least a 3-fold, more preferably at least a 5-fold, higher frequency than the least frequently used codon for the amino acid. For example, human genes use the leucine codon CTG 40% of the time, but use the synonymous CTA only 7% of the time (see Table 2). Thus, CTG is a frequent codon, whereas CTA is a rare codon. Roughly consistent with these frequencies of usage, there are 6 copies in the genome for the gene for the tRNA recognizing CTG, whereas there are only 2 copies of the gene for the tRNA recognizing CTA. Similarly, human genes use the frequent codons TCT and TCC for serine 18% and 22% of the time, respectively, but the rare codon TCG only 5% of the time. TCT and TCC are read, via wobble, by the

same tRNA, which has 10 copies of its gene in the genome, while TCG is read by a tRNA with only 4 copies. It is well known that those mRNAs that are very actively translated are strongly biased to use only the most frequent codons. This includes genes for ribosomal proteins and glycolytic enzymes. On the other hand, mRNAs for relatively non-abundant proteins may use the rare codons.

TABLE 2

Codon usage in <i>Homo sapiens</i> (source: http://www.kazusa.or.jp/codon/)				
Amino Acid	Codon	Number	/1000	Fraction
Gly	GGG	636457.00	16.45	0.25
Gly	GGA	637120.00	16.47	0.25
Gly	GGT	416131.00	10.76	0.16
Gly	GGC	862557.00	22.29	0.34
Glu	GAG	1532589.00	39.61	0.58
Glu	GAA	1116000.00	28.84	0.42
Asp	GAT	842504.00	21.78	0.46
Asp	GAC	973377.00	25.16	0.54
Val	GTG	1091853.00	28.22	0.46
Val	GTA	273515.00	7.07	0.12
Val	GTT	426252.00	11.02	0.18
Val	GTC	562086.00	14.53	0.24
Ala	GCG	286975.00	7.42	0.11
Ala	GCA	614754.00	15.89	0.23
Ala	GCT	715079.00	18.48	0.27
Ala	GCC	1079491.00	27.90	0.40
Arg	AGG	461676.00	11.93	0.21
Arg	AGA	466435.00	12.06	0.21
Ser	AGT	469641.00	12.14	0.15
Ser	AGC	753597.00	19.48	0.24
Lys	AAG	1236148.00	31.95	0.57
Lys	AAA	940312.00	24.30	0.43
Asn	AAT	653566.00	16.89	0.47
Asn	AAC	739007.00	19.10	0.53
Met	ATG	853648.00	22.06	1.00
Ile	ATA	288118.00	7.45	0.17
Ile	ATT	615699.00	15.91	0.36
Ile	ATC	808306.00	20.89	0.47
Thr	ACG	234532.00	6.06	0.11
Thr	ACA	580580.00	15.01	0.28
Thr	ACT	506277.00	13.09	0.25
Thr	ACC	732313.00	18.93	0.36

TABLE 2-continued

Codon usage in <i>Homo sapiens</i> (source: http://www.kazusa.or.jp/codon/)				
Amino Acid	Codon	Number	/1000	Fraction
Trp	TGG	510256.00	13.19	1.00
End	TGA	59528.00	1.54	0.47
Cys	TGT	407020.00	10.52	0.45
Cys	TGC	487907.00	12.61	0.55
End	TAG	30104.00	0.78	0.24
End	TAA	38222.00	0.99	0.30
Tyr	TAT	470083.00	12.15	0.44
Tyr	TAC	592163.00	15.30	0.56
Leu	TTG	498920.00	12.89	0.13
Leu	TTA	294684.00	7.62	0.08
Phe	TTT	676381.00	17.48	0.46
Phe	TTC	789374.00	20.40	0.54
Ser	TCG	171428.00	4.43	0.05
Ser	TCA	471469.00	12.19	0.15
Ser	TCT	585967.00	15.14	0.19
Ser	TCC	684663.00	17.70	0.22
Arg	CGG	443753.00	11.47	0.20
Arg	CGA	239573.00	6.19	0.11
Arg	CGT	176691.00	4.57	0.08
Arg	CGC	405748.00	10.49	0.18
Gln	CAG	1323614.00	34.21	0.74
Gln	CAA	473648.00	12.24	0.26
His	CAT	419726.00	10.85	0.42
His	CAC	583620.00	15.08	0.58
Leu	CTG	1539118.00	39.78	0.40
Leu	CTA	276799.00	7.15	0.07
Leu	CTT	508151.00	13.13	0.13
Leu	CTC	759527.00	19.63	0.20
Pro	CCG	268884.00	6.95	0.11
Pro	CCA	653281.00	16.88	0.28
Pro	CCT	676401.00	17.48	0.29
Pro	CCC	767793.00	19.84	0.32

The propensity for highly expressed genes to use frequent codons is called "codon bias." A gene for a ribosomal protein might use only the 20 to 25 most frequent of the 61 codons, and have a high codon bias (a codon bias close to 1), while a poorly expressed gene might use all 61 codons, and have little or no codon bias (a codon bias close to 0). It is thought that the frequently used codons are codons where larger amounts of the cognate tRNA are expressed, and that

use of these codons allows translation to proceed more rapidly, or more accurately, or both. The PV capsid protein is very actively translated, and has a high codon bias.

Codon Pair Bias

A distinct feature of coding sequences is their codon pair bias. This may be illustrated by considering the amino acid pair Ala-Glu, which can be encoded by 8 different codon pairs. If no factors other than the frequency of each individual codon (as shown in Table 2) are responsible for the frequency of the codon pair, the expected frequency of each of the 8 encodings can be calculated by multiplying the frequencies of the two relevant codons. For example, by this calculation the codon pair GCA-GAA would be expected to occur at a frequency of 0.097 out of all Ala-Glu coding pairs (0.23×0.42 ; based on the frequencies in Table 2). In order to relate the expected (hypothetical) frequency of each codon pair to the actually observed frequency in the human genome the Consensus CDS (CCDS) database of consistently annotated human coding regions, containing a total of 14,795 human genes, was used. This set of genes is the most comprehensive representation of human coding sequences. Using this set of genes the frequencies of codon usage were re-calculated by dividing the number of occurrences of a codon by the number of all synonymous codons coding for the same amino acid. As expected the frequencies correlated closely with previously published ones such as the ones given in Table 2. Slight frequency variations are possibly due to an oversampling effect in the data provided by the codon usage database at Kazusa DNA Research Institute (<http://www.kazusa.or.jp/codon/codon.html>) where 84949 human coding sequences were included in the calculation (far more than the actual number of human genes). The codon frequencies thus calculated were then used to calculate the expected codon-pair frequencies by first multiplying the frequencies of the two relevant codons with each other (see Table 3 expected frequency), and then multiplying this result with the observed frequency (in the entire CCDS data set) with which the amino acid pair encoded by the codon pair in question occurs. In the example of codon pair GCA-GAA, this second calculation gives an expected frequency of 0.098 (compared to 0.97 in the first calculation using the Kazusa dataset). Finally, the actual codon pair frequencies as observed in a set of 14,795 human genes was determined by counting the total number of occurrences of each codon pair in the set and dividing it by the number of all synonymous coding pairs in the set coding for the same amino acid pair (Table 3; observed frequency). Frequency and observed/expected values for the complete set of 3721 (61^2) codon pairs, based on the set of 14,795 human genes, are provided herewith as Supplemental Table 1.

TABLE 3

Codon Pair Scores Exemplified by the Amino Acid Pair Ala-Glu				
amino acid pair	codon pair	expected frequency	observed frequency	obs/exp ratio
AE	GCAGAA	0.098	0.163	1.65
AE	GCAGAG	0.132	0.198	1.51
AE	GCCGAA	0.171	0.031	0.18
AE	GCCGAG	0.229	0.142	0.62
AE	GCGGAA	0.046	0.027	0.57

TABLE 3-continued

Codon Pair Scores Exemplified by the Amino Acid Pair Ala-Glu				
amino acid pair	codon pair	expected frequency	observed frequency	obs/exp ratio
AE	GCGGAG	0.062	0.089	1.44
AE	GCTGAA	0.112	0.145	1.29
AE	GCTGAG	0.150	0.206	1.37
Total		1.000	1.000	

If the ratio of observed frequency/expected frequency of the codon pair is greater than one the codon pair is said to be overrepresented. If the ratio is smaller than one, it is said to be underrepresented. In the example the codon pair GCA-GAA is overrepresented 1.65 fold while the coding pair GCC-GAA is more than 5-fold underrepresented.

Many other codon pairs show very strong bias; some pairs are under-represented, while other pairs are over-represented. For instance, the codon pairs GCCGAA (AlaGlu) and GATCTG (AspLeu) are three- to six-fold under-represented (the preferred pairs being GCAGAG and GACCTG, respectively), while the codon pairs GCCAAG (AlaLys) and AATGAA (AsnGlu) are about two-fold over-represented. It is noteworthy that codon pair bias has nothing to do with the frequency of pairs of amino acids, nor with the frequency of individual codons. For instance, the under-represented pair GATCTG (AspLeu) happens to use the most frequent Leu codon, (CTG).

Codon pair bias was discovered in prokaryotic cells (see Greve et al., 1989), but has since been seen in all other examined species, including humans. The effect has a very high statistical significance, and is certainly not just noise. However, its functional significance, if any, is a mystery. One proposal is that some pairs of tRNAs interact well when they are brought together on the ribosome, while other pairs interact poorly. Since different codons are usually read by different tRNAs, codon pairs might be biased to avoid putting together pairs of incompatible tRNAs (Greve et al., 1989). Another idea is that many (but not all) under-represented pairs have a central CG dinucleotide (e.g., GCCGAA, encoding AlaGlu), and the CG dinucleotide is systematically under-represented in mammals (Buchan et al., 2006; Curran et al., 1995; Fedorov et al., 2002). Thus, the effects of codon pair bias could be of two kinds—one an indirect effect of the under-representation of CG in the mammalian genome, and the other having to do with the efficiency, speed and/or accuracy of translation. It is emphasized that the present invention is not limited to any particular molecular mechanism underlying codon pair bias.

As discussed more fully below, codon pair bias takes into account the score for each codon pair in a coding sequence averaged over the entire length of the coding sequence. According to the invention, codon pair bias is determined by

$$CPB = \sum_{i=1}^k \frac{CPS_i}{k-1}.$$

Accordingly, similar codon pair bias for a coding sequence can be obtained, for example, by minimized codon

pair scores over a subsequence or moderately diminished codon pair scores over the full length of the coding sequence.

Since all 61 sense codons and all sense codon pairs can certainly be used, it would not be expected that substituting a single rare codon for a frequent codon, or a rare codon pair for a frequent codon pair, would have much effect. Therefore, many previous investigations of codon and codon pair bias have been done via informatics, not experimentation.

10 One investigation of codon pair bias that was based on experimental work was the study of Irwin et al. (1995), who found, counterintuitively, that certain over-represented codon pairs caused slower translation. However, this result could not be reproduced by a second group (Cheng and Goldman, 2001), and is also in conflict with results reported below. Thus, the present results (see below) may be the first experimental evidence for a functional role of codon pair bias.

Certain experiments disclosed herein relate to re-coding 20 viral genome sequences, such as the entire capsid region of PV, involving around 1000 codons, to separately incorporate both poor codon bias and poor codon pair bias into the genome. The rationale underlying these experiments is that if each substitution creates a small effect, then all substitutions together should create a large effect. Indeed, it turns out that both deoptimized codon bias, and deoptimized codon pair bias, separately create non-viable viruses. As discussed in more detail in the Examples, preliminary data suggest that inefficient translation is the major mechanism for reducing 30 the viability of a virus with poor codon bias or codon pair bias. Irrespective of the precise mechanism, the data indicate that the large-scale substitution of synonymous deoptimized codons into a viral genome results in severely attenuated viruses. This procedure for producing attenuated viruses has 35 been dubbed SAVE (Synthetic Attenuated Virus Engineering).

According to the invention, viral attenuation can be 40 accomplished by changes in codon pair bias as well as codon bias. However, it is expected that adjusting codon pair bias is particularly advantageous. For example, attenuating a virus through codon bias generally requires elimination of common codons, and so the complexity of the nucleotide sequence is reduced. In contrast, codon pair bias reduction or minimization can be accomplished while maintaining far 45 greater sequence diversity, and consequently greater control over nucleic acid secondary structure, annealing temperature, and other physical and biochemical properties. The work disclosed herein includes attenuated codon pair bias-reduced or -minimized sequences in which codons are 50 shuffled, but the codon usage profile is unchanged.

Viral attenuation can be confirmed in ways that are well known to one of ordinary skill in the art. Non-limiting examples induce plaque assays, growth measurements, and reduced lethality in test animals. The instant application 55 demonstrates that the attenuated viruses are capable of inducing protective immune responses in a host.

Synthetic Attenuated Virus Engineering (SAVE)

SAVE employs specifically designed computer software and modern methods of nucleic acid synthesis and assembly 60 to re-code and re-synthesize the genomes of viruses. This strategy provides an efficient method of producing vaccines against various medically important viruses for which efficacious vaccines are sought.

Two effective polio vaccines, an inactivated polio vaccine 65 (IPV) developed by Jonas Salk and an oral polio vaccine (OPV) comprising live attenuated virus developed by Albert Sabin, respectively, have been available sine the 1950's.

Indeed, a global effort to eradicate poliomyelitis, begun in 1988 and led by the World Health Organization (WHO), has succeeded in eradicating polio from most of the countries in the world. The number of annual diagnosed cases has been reduced from the hundreds of thousands to less than two thousand in 2005, occurring mainly in India and in Nigeria. However, a concern regarding the wide use of the OPV is that it can revert to a virulent form, and though believed to be a rare event, outbreaks of vaccine-derived polio have been reported (Georgescu et al., 1997; Kew et al., 2002; Shimizu et al., 2004). In fact, as long as the live poliovirus vaccine strains are used, each carrying less than 7 attenuating mutations, there is a possibility that this strain will revert to wt, and such reversion poses a serious threat to the complete eradication of polio. Thus, the WHO may well need a new polio vaccine to combat the potential of reversion in the closing stages of its efforts at polio eradication, and this provides one rationale for the studies disclosed herein on the application of SAVE to PV. However, PV was selected primarily because it is an excellent model system for developing SAVE.

During re-coding, essential nucleic acid signals in the viral genome are preserved, but the efficiency of protein translation is systematically reduced by deoptimizing codon bias, codon pair bias, and other parameters such as RNA secondary structure and CpG dinucleotide content, C+G content, translation frameshift sites, translation pause sites, or any combination thereof. This deoptimization may involve hundreds or thousands of changes, each with a small effect. Generally, deoptimization is performed to a point at which the virus can still be grown in some cell lines (including lines specifically engineered to be permissive for a particular virus), but where the virus is avirulent in a normal animal or human. Such avirulent viruses are excellent candidates for either a killed or live vaccine since they encode exactly the same proteins as the fully virulent virus and accordingly provoke exactly the same immune response as the fully virulent virus. In addition, the SAVE process offers the prospect for fine tuning the level of attenuation; that is, it provides the capacity to design synthetic viruses that are deoptimized to a roughly predictable extent. Design, synthesis, and production of viral particles is achievable in a timeframe of weeks once the genome sequence is known, which has important advantages for the production of vaccines in potential emergencies. Furthermore, the attenuated viruses are expected to have virtually no potential to revert to virulence because of the extremely large numbers of deleterious nucleotide changes involved. This method may be generally applicable to a wide range of viruses, requiring only knowledge of the viral genome sequence and a reverse genetics system for any particular virus.

Viral Attenuation by Deoptimizing Codon Bias

If one uses the IC₅₀-ratio of control cells/test cells method as described above, then compounds with CSG values less than or equal to 1 would not generally be considered to be good clinical candidate compounds, whereas compounds with CSG values of greater than approximately 10 would be quite promising and worthy of further consideration.

As a means of engineering attenuated viruses, the capsid coding region of poliovirus type 1 Mahoney [PV(M)] was re-engineered by making changes in synonymous codon usage. The capsid region comprises about a third of the virus and is very actively translated. One mutant virus (virus PV-AB), having a very low codon bias due to replacement of the largest possible number of frequently used codons with rare synonymous codons was created. As a control, another virus (PV-SD) was created having the largest pos-

sible number of synonymous codon changes while maintaining the original codon bias. See FIGS. 1 and 2. Thus, PV-SD is a virus having essentially the same codons as the wt, but in shuffled position while encoding exactly the same proteins. In PV-SD, no attempt was made to increase or reduce codon pair bias by the shuffling procedure. See Example 1. Despite 934 nucleotide changes in the capsid-coding region, PV-SD RNA produced virus with characteristics indistinguishable from wt. In contrast, no viable virus was recovered from PV-AB carrying 680 silent mutations. See Example 2.

A trivial explanation of the inviability of PV-AB is that just one of the nucleotide changes is somehow lethal, while the other 679 are harmless. For instance, a nucleotide change could be lethal for some catastrophic but unappreciated reason, such as preventing replication. This explanation is unlikely, however. Although PV does contain important regulatory elements in its RNA, such as the CRE, it is known that no such elements exist inside the capsid coding region. This is supported by the demonstration that the entire capsid coding region can be deleted without affecting normal replication of the residual genome within the cell, though of course viral particles cannot be formed (Kaplan and Racamiello, 1988).

To address questions concerning the inviability of certain re-engineered viruses, sub-segments of the capsid region of virus PV-AB were subcloned into the wild type virus. See Example 1 and FIG. 3. Incorporating large subcloned segments (including non-overlapping segments) proved lethal, while small subcloned segments produced viable (with one exception) but sick viruses. "Sickness" is revealed by many assays: for example, segments of poor codon bias cause poor titers (FIG. 3B) and small plaques (FIGS. 3C-H). It is particularly instructive that in general, large, lethal segments can be divided into two sub-segments, both of which are alive but sick (FIG. 3). These results rule out the hypothesis that inviability is due to just one change; instead, at minimum, many changes must be contributing to the phenotype.

There is an exceptional segment from position 1513 to 40 2470. This segment is fairly small, but its inclusion in the PV genome causes inviability. It is not known at present whether or not this fragment can be subdivided into subfragments that merely cause sickness and do not inactivate the virus. It is conceivable that this segment does contain a highly deleterious change, possibly a translation frameshift site.

Since poor codon bias naturally suggests an effect on translation, translation of the proteins encoded by virus PV-AB was tested. See Example 5 and FIG. 5. Indeed, all the sick viruses translated capsid protein poorly (FIG. 5B). Translation was less efficient in the sicker viruses, consistent with poor translation being the cause of the sickness. Translation was improved essentially to wt levels in reactions that were supplemented with excess tRNAs and amino acids (FIG. 5A), consistent with the rate of recognition of rare codons being limiting.

As a second test of whether deoptimized codon bias was causing inefficient translation, portions of wt and deoptimized capsid were fused to the N-terminus of firefly luciferase in a dicistronic reporter construct. See Example 5 and FIG. 6. In these fusion constructs, translation of luciferase depends on translation of the N-terminally fused capsid protein. Again, it was found that translation of the capsid proteins with deoptimized codons was poor, and was worse in the sicker viruses, suggesting a cause-and-effect relationship. Thus, the data suggest that the hundreds of rare codons in the PV-AB virus cause inviability largely because of poor translation. Further, the poor translation seen in vitro

41

and the viral sickness seen in cultured cells are also reflected in infections of animals. Even for one of the least debilitated deoptimized viruses, PV-AB²⁴⁷⁰⁻²⁹⁵⁴, the number of viral particles needed to cause disease in mice was increased by about 100-fold. See Example 4, Table 4.

Burns et al. (2006) have recently described some similar experiments with the Sabin type 2 vaccine strain of PV and reached similar conclusions. Burns et al. synthesized a completely different codon-deoptimized virus (i.e., the nucleotide sequences of the PV-AB virus described herein and their “abcd” virus are very different), and yet got a similar degree of debilitation using similar assays. Burns et al. did not test their viral constructs in host organisms for attenuation. However, their result substantiates the view that SAVI is predictable, and that the results are not greatly dependent on the exact nucleotide sequence.

Viral Attenuation by Deoptimizing Codon Pair Bias

According to the invention, codon pair bias can be altered independently of codon usage. For example, in a protein encoding sequence of interest, codon pair bias can be altered simply by directed rearrangement of its codons. In particular, the same codons that appear in the parent sequence, which can be of varying frequency in the host organism, are used in the altered sequence, but in different positions. In the simplest form, because the same codons are used as in the parent sequence, codon usage over the protein coding region being considered remains unchanged (as does the encoded amino acid sequence). Nevertheless, certain codons appear in new contexts, that is, preceded by and/or followed by codons that encode the same amino acid as in the parent sequence, but employing a different nucleotide triplet. Ideally, the rearrangement of codons results in codon pairs that are less frequent than in the parent sequence. In practice, rearranging codons often results in a less frequent codon pair at one location and a more frequent pair at a second location. By judicious rearrangement of codons, the codon pair usage bias over a given length of coding sequence can be reduced relative to the parent sequence. Alternatively, the codons could be rearranged so as to produce a sequence that makes use of codon pairs which are more frequent in the host than in the parent sequence.

Codon pair bias is evaluated by considering each codon pair in turn, scoring each pair according to the frequency that the codon pair is observed in protein coding sequences of the host, and then determining the codon pair bias for the sequence, as disclosed herein. It will be appreciated that one can create many different sequences that have the same codon pair bias. Also, codon pair bias can be altered to a greater or lesser extent, depending on the way in which codons are rearranged. The codon pair bias of a coding sequence can be altered by recoding the entire coding sequence, or by recoding one or more subsequences. As used herein, “codon pair bias” is evaluated over the length of a coding sequence, even though only a portion of the sequence may be mutated. Because codon pairs are scored in the context of codon usage of the host organism, a codon pair bias value can be assigned to wild type viral sequences and mutant viral sequences. According to the invention, a virus can be attenuated by recoding all or portions of the protein encoding sequences of the virus so as to reduce its codon pair bias.

According to the invention, codon pair bias is a quantitative property determined from codon pair usage of a host. Accordingly, absolute codon pair bias values may be determined for any given viral protein coding sequence. Alternatively, relative changes in codon pair bias values can be determined that relate a deoptimized viral protein coding

42

sequence to a “parent” sequence from which it is derived. As viruses come in a variety of types (i.e., types I to VII by the Baltimore classification), and natural (i.e., virulent) isolates of different viruses yield different values of absolute codon pair bias, it is relative changes in codon pair bias that are usually more relevant to determining desired levels of attenuation. Accordingly, the invention provides attenuated viruses and methods of making such, wherein the attenuated viruses comprise viral genomes in which one or more protein encoding nucleotide sequences have codon pair bias reduced by mutation. In viruses that encode only a single protein (i.e., a polyprotein), all or part of the polyprotein can be mutated to a desired degree to reduce codon pair bias, and all or a portion of the mutated sequence can be provided in a recombinant viral construct. For a virus that separately encodes multiple proteins, one can reduce the codon pair bias of all of the protein encoding sequences simultaneously, or select only one or a few of the protein encoding sequences for modification. The reduction in codon pair bias is determined over the length of a protein encoding sequences, and is at least about 0.05, or at least about 0.1, or at least about 0.15, or at least about 0.2, or at least about 0.3, or at least about 0.4. Depending on the virus, the absolute codon pair bias, based on codon pair usage of the host, can be about -0.05 or less, or about -0.1 or less, or about -0.15 or less, or about -0.2 or less, or about -0.3 or less, or about -0.4 or less.

It will be apparent that codon pair bias can also be superimposed on other sequence variation. For example, a coding sequence can be altered both to encode a protein or polypeptide which contains one or more amino acid changes and also to have an altered codon pair bias. Also, in some cases, one may shuffle codons to maintain exactly the same codon usage profile in a codon-bias reduced protein encoding sequence as in a parent protein encoding sequence. This procedure highlights the power of codon pair bias changes, but need not be adhered to. Alternatively, codon selection can result in an overall change in codon usage in a coding sequence. In this regard, it is noted that in certain examples provided herein, (e.g., the design of PV-Min) even if the codon usage profile is not changed in the process of generating a codon pair bias minimized sequence, when a portion of that sequence is subcloned into an unmutated sequence (e.g., PV-MinXY or PV-MinZ), the codon usage profile over the subcloned portion, and in the hybrid produced, will not match the profile of the original unmutated protein coding sequence. However, these changes in codon usage profile have minimal effect of codon pair bias.

Similarly, it is noted that, by itself, changing a nucleotide sequence to encode a protein or polypeptide with one or many amino acid substitutions is also highly unlikely to produce a sequence with a significant change in codon pair bias. Consequently, codon pair bias alterations can be recognized even in nucleotide sequences that have been further modified to encode a mutated amino acid sequence. It is also noteworthy that mutations meant by themselves to increase codon bias are not likely to have more than a small effect on codon pair bias. For example, as disclosed herein, the codon pair bias for a poliovirus mutant recoded to maximize the use of nonpreferred codons (PV-AB) is decreased from wild type (PV-1(M)) by only about 0.05. Also noteworthy is that such a protein encoding sequence have greatly diminished sequence diversity. To the contrary, substantial sequence diversity is maintained in codon pair bias modified sequences of the invention. Moreover, the significant reduction in codon pair bias obtainable without increased use of rare codons suggests that instead of maximizing the use of

nonpreferred codons, as in PV-AB, it would be beneficial to rearrange nonpreferred codons with a sufficient number of preferred codons in order to more effectively reduce codon pair bias.

The extent and intensity of mutation can be varied depending on the length of the protein encoding nucleic acid, whether all or a portion can be mutated, and the desired reduction of codon pair bias. In an embodiment of the invention, a protein encoding sequence is modified over a length of at least about 100 nucleotide, or at least about 200 nucleotides, or at least about 300 nucleotides, or at least about 500 nucleotides, or at least about 1000 nucleotides.

As discussed above, the term “parent” virus or “parent” protein encoding sequence is used herein to refer to viral genomes and protein encoding sequences from which new sequences, which may be more or less attenuated, are derived. Accordingly, a parent virus can be a “wild type” or “naturally occurring” prototypes or isolate or variant or a mutant specifically created or selected on the basis of real or perceived desirable properties.

Using de novo DNA synthesis, the capsid coding region (the P1 region from nucleotide 755 to nucleotide 3385) of PV(M) was redesigned to introduce the largest possible number of rarely used codon pairs (virus PV-Min) (SEQ ID NO:4) or the largest possible number of frequently used codon pairs (virus PV-Max) (SEQ ID NO:5), while preserving the codon bias of the wild type virus. See Example 7. That is, the designed sequences use the same codons as the parent sequence, but they appear in a different order. The PV-Max virus exhibited one-step growth kinetics and killing of infected cells essentially identical to wild type virus. (That growth kinetics are not increased for a codon pair maximized virus relative to wild type appears to hold true for other viruses as well.) Conversely, cells transfected with PV-Min mutant RNA were not killed, and no viable virus could be recovered. Subcloning of fragments (PV-Min⁷⁵⁵⁻²⁴⁷⁰, PV-Min₂₄₇₀₋₃₃₈₆) of the capsid region of PV-Min into the wt background produced very debilitated, but not dead, virus. See Example 7 and FIG. 8. This result substantiates the hypothesis that deleterious codon changes are preferably widely distributed and demonstrates the simplicity and effectiveness of varying the extent of the codon pair deoptimized sequence that is substituted into a wild type parent virus genome in order to vary the codon pair bias for the overall sequence and the attenuation of the viral product. As seen with PV-AB viruses, the phenotype of PV-Min viruses is a result of reduced specific infectivity of the viral particles rather than of lower production of progeny virus.

Virus with deoptimized codon pair bias are attenuated. As exemplified below, (see Example 8, and Table 5), CD155tg mice survived challenge by intracerebral injection of attenuated virus in amounts 1000-fold higher than would be lethal for wild type virus. These findings demonstrate the power of deoptimization of codon pair bias to minimize lethality of a virus. Further, the viability of the virus can be balanced with a reduction of infectivity by choosing the degree of codon pair bias deoptimization. Further, once a degree or ranges of degrees of codon pair bias deoptimization is determined that provides desired attenuation properties, additional sequences can be designed to attain that degree of codon pair bias. For example, SEQ ID NO:6 provides a poliovirus sequence with a codon pair bias of about -0.2, and mutations distributed over the region encompassing the mutated portions of PV-MinXY and PV-MinZ (i.e., PV⁷⁵⁵⁻³³⁸⁵).

Algorithms for Sequence Design

The inventors have developed several novel algorithms for gene design that optimize the DNA sequence for par-

ticular desired properties while simultaneously coding for the given amino acid sequence. In particular, algorithms for maximizing or minimizing the desired RNA secondary structure in the sequence (Cohen and Skiena, 2003) as well as maximally adding and/or removing specified sets of patterns (Skiena, 2001), have been developed. The former issue arises in designing viable viruses, while the latter is useful to optimally insert restriction sites for technological reasons. The extent to which overlapping genes can be designed that simultaneously encode two or more genes in alternate reading frames has also been studied (Wang et al., 2006). This property of different functional polypeptides being encoded in different reading frames of a single nucleic acid is common in viruses and can be exploited for technological purposes such as weaving in antibiotic resistance genes.

The first generation of design tools for synthetic biology has been built, as described by Jayaraj et al. (2005) and Richardson et al. (2006). These focus primarily on optimizing designs for manufacturability (i.e., oligonucleotides without local secondary structures and end repeats) instead of optimizing sequences for biological activity. These first-generation tools may be viewed as analogous to the early VLSI CAD tools built around design rule-checking, instead of supporting higher-order design principles.

As exemplified herein, a computer-based algorithm can be used to manipulate the codon pair bias of any coding region. The algorithm has the ability to shuffle existing codons and to evaluate the resulting CPB, and then to reshuffle the sequence, optionally locking in particularly “valuable” codon pairs. The algorithm also employs a form of “simulated annealing” so as not to get stuck in local minima. Other parameters, such as the free energy of folding of RNA, may optional be under the control of the algorithm as well, in order to avoid creation of undesired secondary structures. The algorithm can be used to find a sequence with a minimum codon pair bias, and in the event that such a sequence does not provide a viable virus, the algorithm can be adjusted to find sequences with reduced, but not minimized biases. Of course, a viable viral sequence could also be produced using only a subsequence of the computer mimimized sequence.

Whether or not performed with the aid of a computer, using, for example, a gradient descent, or simulated annealing, or other minimization routine. An example of the procedure that rearranges codons present in a starting sequence can be represented by the following steps:

- 50 1) Obtain wildtype viral genome sequence.
- 2) Select protein coding sequences to target for attenuated design.
- 3) Lock down known or conjectured DNA segments with non-coding functions.
- 55 4) Select desired codon distribution for remaining amino acids in redesigned proteins.
- 5) Perform random shuffle of unlocked codon positions and calculate codon-pair score.
- 6) Further reduce (or increase) codon-pair score optionally employing a simulated annealing procedure.
- 7) Inspect resulting design for excessive secondary structure and unwanted restriction site:
if yes → go to step (5) or correct the design by replacing problematic regions with wildtype sequences and go to step (8).
8. Synthesize DNA sequence corresponding to virus design.

45

9. Create viral construct and assess expression:
if too attenuated, prepare subclone construct and goto 9;
if insufficiently attenuated, goto 2.

Source code (PERL script) of a computer based simulated annealing routine is provided.

Alternatively, one can devise a procedure which allows each pair of amino acids to be deoptimized by choosing a codon pair without a requirement that the codons be swapped out from elsewhere in the protein encoding sequence.

Molecular Mechanisms of Viral Attenuation: Characterization of Attenuated PV using High-Throughput Methods

As described above and in greater detail in the Examples, two synthetic, attenuated polioviruses encoding exactly the same proteins as the wildtype virus, but having altered codon bias or altered codon pair bias, were constructed. One virus uses deoptimized codons; the other virus uses deoptimized codon pairs. Each virus has many hundreds of nucleotide changes with respect to the wt virus.

The data presented herein suggest that these viruses are attenuated because of poor translation. This finding, if correct, has important consequences. First, the reduced fitness/virulence of each virus is due to small defects at hundreds of positions spread over the genome. Thus, there is essentially no chance of the virus reverting to wildtype, and so the virus is a good starting point for either a live or killed vaccine. Second, if the reduced fitness/virulence is due to additive effects of hundreds of small defects in translation, this method of reducing fitness with minimal risk of reversion should be applicable to many other viruses.

Though it is emphasized that the present invention is not limited to any particular mode of operation or underlying molecular mechanism, ongoing studies are aimed at distinguishing these alternative hypotheses. The ongoing investigations involve use of high throughput methods to scan through the genomes of various attenuated virus designs such as codon and codon pair deoptimized poliovirus and influenza virus, and to construct chimeras by placing overlapping 300-bp portions of each mutant virus into a wt context. See Example 11. The function of these chimeric viruses are then assayed. A finding that most chimeras are slightly, but not drastically, less fit than wild type, as suggested by the preliminary data disclosed herein, corroborates the “incremental loss of function” hypothesis, wherein many deleterious mutations are distributed throughout the regions covered by the chimeras. Conversely, a finding that most of the chimeras are similar or identical to wt, whereas one or only a few chimeras are attenuated like the parental mutant, suggests that there are relatively few positions in the sequence where mutation results in attenuation and that attenuation at those positions is significant.

As described in Example 12, experiments are performed to determine how codon and codon-pair deoptimization affect RNA stability and abundance, and to pinpoint the parameters that impair translation of the re-engineered viral genome. An understanding of the molecular basis of this impairment will further enhance the applicability of the SAVE approach to a broad range of viruses. Another conceivable mechanism underlying translation impairment is translational frameshifting, wherein the ribosome begins to translate a different reading frame, generating a spurious, typically truncated polypeptide up to the point where it encounters an in-frame stop codon. The PV genomes carrying the AB mutant segment from residue 1513 to 2470 are not only non-viable, but also produce a novel protein band during in vitro translation of approximately 42-44 kDa (see FIG. 5A). The ability of this AB¹⁵¹³⁻²⁴⁷⁰ fragment to inac-

46

tivate PV, as well as its ability to induce production of the novel protein, may reflect the occurrence of a frameshift event and this possibility is also being investigated. A filter for avoiding the introduction of frameshifting sites is built into the SAVE design software.

More detailed investigations of translational defects are conducted using various techniques including, but not limited to, polysome profiling, toeprinting, and luciferase assays of fusion proteins, as described in Example 12.

10 Molecular Biology of Poliovirus

While studies are ongoing to unravel the mechanisms underlying viral attenuation by SAVE, large-scale codon deoptimization of the PV capsid coding region revealed interesting insights into the biology of PV itself. What determines the PFU/particle ratio (specific infectivity) of a virus has been a longstanding question. In general, failure at any step during the infectious life cycle before the establishment of a productive infection will lead to an abortive infection and, therefore, to the demise of the infecting particle. In the case of PV, it has been shown that approximately 100 virions are required to result in one infectious event in cultured cells (Joklik and Darnell, 1961; Schwerdt and Fogh, 1957). That is, of 100 particles inoculated, only approximately one is likely to successfully complete all steps at the level of receptor binding (step 1), followed by internalization and uncoating (step 2), initiation of genome translation (step 3), polyprotein translation (step 4), RNA replication (step 5), and encapsidation of progeny (step 6).

In the infectious cycle of AB-type viruses described here, steps 1 and 2 should be identical to a PV(M) infection as their capsids are identical. Likewise, identical 5' nontranslated regions should perform equally well in assembly of a translation complex (step 3). Viral polyprotein translation, on the other hand (step 4), is severely debilitated due to the introduction of a great number of suboptimal synonymous codons in the capsid region (FIGS. 5 and 6). It is thought that the repeated encounter of rare codons by the translational machinery causes stalling of the ribosome as, by the laws of mass action, rare aminoacyl-tRNA will take longer to diffuse into the A site on the ribosome. As peptide elongation to a large extent is driven by the concentration of available aminoacyl-tRNA, dependence of an mRNA on many rare tRNAs consequently lengthens the time of translation (Gustafsson et al., 2004). Alternatively, excessive stalling of the ribosome may cause premature dissociation of the translation complex from the RNA and result in a truncated protein destined for degradation. Both processes lead to a lower protein synthesis rate per mRNA molecule per unit of time. While the data presented herein suggest that the phenotypes of codon-deoptimized viruses are determined by the rate of genome translation, other mechanistic explanations may be possible. For example, it has been suggested that the conserved positions of rare synonymous codons throughout the viral capsid sequence in Hepatitis A virus are of functional importance for the proper folding of the nascent polypeptide by introducing necessary translation pauses (Sánchez et al., 2003). Accordingly, large-scale alteration of the codon composition may conceivably change some of these pause sites to result in an increase of misfolded capsid proteins.

Whether these considerations also apply to the PV capsid is not clear. If so, an altered phenotype would have been expected with the PV-SD design, in which the wt codons were preserved, but their positions throughout the capsid were completely changed. That is, none of the purported pause sites would be at the appropriate position with respect

to the protein sequence. No change in phenotype, however, was observed and PV-SD translated and replicated at wild type levels (FIG. 3B).

Another possibility is that the large-scale codon alterations in the tested designs may create fortuitous dominant-negative RNA elements, such as stable secondary structures, or sequences that may undergo disruptive long-range interactions with other regions of the genome.

It is assumed that all steps prior to, and including, virus uncoating should be unchanged when wt and the mutant viruses, described herein are compared. This is supported by the observation that the eclipse period for all these isolates is similar (FIG. 3B). The dramatic reduction in PFU/particle ratio is, therefore, likely to be a result of the reduced translation capacity of the deoptimized genomes, i.e., the handicap of the mutant viruses is determined intracellularly.

It is generally assumed that the relatively low PFU/particle ratio of picornaviruses of 1/100 to 1/1,000 (Rueckert, 1985) is mainly determined by structural alterations at the receptor binding step, either prior to or at the level of cell entry. The formation of 13S particles that are hardly infectious may be the major culprit behind the inefficiency of poliovirus infectivity (Hogle, 2002). However, certain virus mutants seem to sidestep A particle conversion without resulting in a higher specific infectivity, an observation suggesting that other post-entry mechanisms may be responsible for the low PFU/particle ratio (Dove and Racaniello, 1997).

The present data provide clear evidence for such post-entry interactions between virus and cell, and suggest that these, and not pre-entry events, contribute to the distinct PFU/particle ratio of poliovirus. As all replication proteins in poliovirus are located downstream of P1 on the polyprotein, they critically depend upon successful completion of P1 translation. Lowering the rate of P1 translation therefore lowers translation of all replication proteins to the same extent. This, in turn, likely leads to a reduced capacity of the virus to make the necessary modifications to the host cell required for establishment of a productive infection, such as shutdown of host cell translation or prevention of host cell innate responses. While codon deoptimization, as described herein, is likely to effect translation at the peptide elongation step, reduced initiation of translation can also be a powerful attenuating determinant as well, as has been shown for mutations in the internal ribosomal entry site in the Sabin vaccine strains of poliovirus (Svitkin et al., 1993; 1985).

On the basis of these considerations, it is predicted that many mutant phenotypes attributable to defects in genome translation or early genome replication actually manifest themselves by lowering PFU/particle ratios. This would be the case as long as the defect results in an increased chance of abortive infection. Since in almost all studies the omnipresent plaque assay is the virus detection method of choice, a reduction in the apparent virus titer is often equated with a reduction in virus production per se. This may be an inherent pitfall that can be excused with the difficulties of characterizing virus properties at the single-cell level. Instead, most assays are done on a large population of cells. A lower readout of the chosen test (protein synthesis, RNA replication, virus production as measured in PFU) is taken at face value as an indicator of lower production on a per-cell basis, without considering that virus production in a cell may be normal while the number of cells producing virus is reduced.

The near-identical production of particles per cell by codon-deoptimized viruses indicates that the total of protein produced after extended period of times is not severely

affected, whereas the rate of protein production has been drastically reduced. This is reflected in the delayed appearance of CPE, which may be a sign that the virus has to go through more RNA replication cycles to build up similar intracellular virus protein concentrations. It appears that codon-deoptimized viruses are severely handicapped in establishing a productive infection because the early translation rate of the incoming infecting genome is reduced. As a result of this lower translation rate, PV proteins essential for disabling the cell's antiviral responses (most likely proteinases 2A^{pro} and 3C^{pro}) are not synthesized at sufficient amounts to pass this crucial hurdle in the life cycle quickly enough. Consequently, there is a better chance for the cell to eliminate the infection before viral replication could unfold and take over the cell. Thus, the likelihood for productive infection events is reduced and the rate of abortive infection is increased. However, in the case where a codon-deoptimized virus does succeed in disabling the cell, this virus will produce nearly identical amounts of progeny to the wild type. The present data suggest that a fundamental difference may exist between early translation (from the incoming RNA genome) and late translation during the replicative phase, when the cell's own translation is largely shut down. Although this may be a general phenomenon, it might be especially important in the case of codon-deoptimized genomes. Host cell shutdown very likely results in an overabundance of free aminoacyl-tRNAs, which may overcome the imposed effect of the suboptimal codon usage as the PV genomes no longer have to compete with cellular RNAs for translation resources. This, in fact, may be analogous to observations with the modified in vitro translation system described herein (FIG. 5B). Using a translation extract that was not nuclease-treated (and thus contained cellular mRNAs) and not supplemented with exogenous amino acids or tRNAs, clear differences were observed in the translation capacity of different capsid design mutants. Under these conditions, viral genomes have to compete with cellular mRNAs in an environment where supplies are limited. In contrast, in the traditional translation extract, in which endogenous mRNAs were removed and excess tRNAs and amino acids were added, all PV RNAs translated equally well regardless of codon bias (FIG. 5A). These two different in vitro conditions may be analogous to in vivo translation during the early and late phases in the PV-infected cell.

One key finding of the present study is the realization that, besides the steps during the physical interaction and uptake of virus, the PFU/particle ratio also largely reflects the virus' capacity to overcome host cell antiviral responses. This suggests that picornaviruses are actually quite inefficient in winning this struggle, and appear to have taken the path of evolving small genomes that can quickly replicate before the cell can effectively respond. As the data show, slowing down translation rates by only 30% in PV-AB²⁴⁷⁰⁻²⁹⁵⁴ (see FIG. 6) leads to a 1,000-fold higher rate of abortive infection as reflected in the lower specific infectivity (FIG. 4D). Picornaviruses apparently not only replicate at the threshold of error catastrophe (Crotty et al., 2001; Holland et al., 1990) but also at the threshold of elimination by the host cell's antiviral defenses. This effect may have profound consequences for the pathogenic phenotype of a picornavirus. The cellular antiviral processes responsible for the increased rate of aborted infections by codon-deoptimized viruses are not completely understood at present. PV has been shown to both induce and inhibit apoptosis (Belov et al., 2003; Girard et al., 1999; Tolskaya et al., 1995). Similarly PV interferes with the interferon pathway by cleaving NF-κB (Neznanov et al., 2005). It is plausible that a PV with a reduced rate of

early genome translation still induces antiviral responses in the same way as a wt virus (induction of apoptosis and interferon by default) but then, due to low protein synthesis, has a reduced potential of inhibiting these processes. This scenario would increase the likelihood of the cell aborting a nascent infection and could explain the observed phenomena. At the individual cell level, PV infection is likely to be an all-or-nothing phenomenon. Viral protein and RNA syntheses likely need to be within a very close to maximal range in order to ensure productive infection.

Attenuated Virus Vaccine Compositions

The present invention provides a vaccine composition for inducing a protective immune response in a subject comprising any of the attenuated viruses described herein and a pharmaceutically acceptable carrier.

It should be understood that an attenuated virus of the invention, where used to elicit a protective immune response in a subject or to prevent a subject from becoming afflicted with a virus-associated disease, is administered to the subject in the form of a composition additionally comprising a pharmaceutically acceptable carrier. Pharmaceutically acceptable carriers are well known to those skilled in the art and include, but are not limited to, one or more of 0.01-0.1M and preferably 0.05M phosphate buffer, phosphate-buffered saline (PBS), or 0.9% saline. Such carriers also include aqueous or non-aqueous solutions, suspensions, and emulsions. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, saline and buffered media. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's and fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers such as those based on Ringer's dextrose, and the like. Solid compositions may comprise nontoxic solid carriers such as, for example, glucose, sucrose, mannitol, sorbitol, lactose, starch, magnesium stearate, cellulose or cellulose derivatives, sodium carbonate and magnesium carbonate. For administration in an aerosol, such as for pulmonary and/or intranasal delivery, an agent or composition is preferably formulated with a nontoxic surfactant, for example, esters or partial esters of C6 to C22 fatty acids or natural glycerides, and a propellant. Additional carriers such as lecithin may be included to facilitate intranasal delivery. Pharmaceutically acceptable carriers can further comprise minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives and other additives, such as, for example, antimicrobials, antioxidants and chelating agents, which enhance the shelf life and/or effectiveness of the active ingredients. The instant compositions can, as is well known in the art, be formulated so as to provide quick, sustained or delayed release of the active ingredient after administration to a subject.

In various embodiments of the instant vaccine composition, the attenuated virus (i) does not substantially alter the synthesis and processing of viral proteins in an infected cell; (ii) produces similar amounts of virions per infected cell as wt virus; and/or (iii) exhibits substantially lower virion-specific infectivity than wt virus. In further embodiments, the attenuated virus induces a substantially similar immune response in a host animal as the corresponding wt virus.

This invention also provides a modified host cell line specially isolated or engineered to be permissive for an attenuated virus that is inviable in a wild type host cell. Since the attenuated virus cannot grow in normal (wild type) host

cells, it is absolutely dependent on the specific helper cell line for growth. This provides a very high level of safety for the generation of virus for vaccine production. Various embodiments of the instant modified cell line permit the growth of an attenuated virus, wherein the genome of said cell line has been altered to increase the number of genes encoding rare tRNAs.

In preferred embodiments, the rare codons are CTA (coding for Leu), TCG (Ser), and CCG (Pro). In different embodiments, one, two, or all three of these rare codons are substituted for synonymous frequent codons in the viral genome. For example, all Leu codons in the virus may be changed to CTA; all Ser codons may be changed to TCG; all Pro codons may be changed to CCG; the Leu and Ser, or Leu and Pro, or Ser and Pro codons may be replaced by the identified rare codons; or all Leu, Ser, and Pro codons may be changed to CTA, TCG, and CCG, respectively, in a single virus. Further, a fraction of the relevant codons, i.e., less than 100%, may be changed to the rare codons. Thus, the proportion of codons substituted may be about 20%, 40%, 60%, 80% or 100% of the total number of codons.

In certain embodiments, these substitutions are made only in the capsid region of the virus, where a high rate of translation is most important. In other embodiments, the substitutions are made throughout the virus. In further embodiments, the cell line over-expresses tRNAs that bind to the rare codons.

This invention further provides a method of synthesizing any of the attenuated viruses described herein, the method comprising (a) identifying codons in multiple locations within at least one non-regulatory portion of the viral genome, which codons can be replaced by synonymous codons; (b) selecting a synonymous codon to be substituted for each of the identified codons; and (c) substituting a synonymous codon for each of the identified codons.

In certain embodiments of the instant methods, steps (a) and (b) are guided by a computer-based algorithm for Synthetic Attenuated Virus Engineering (SAVE) that permits design of a viral genome by varying specified pattern sets of deoptimized codon distribution and/or deoptimized codon-pair distribution within preferred limits. The invention also provides a method wherein, the pattern sets alternatively or additionally comprise, density of deoptimized codons and deoptimized codon pairs, RNA secondary structure, CpG dinucleotide content, C+G content, overlapping coding frames, restriction site distribution, frameshift sites, or any combination thereof.

In other embodiments, step (c) is achieved by de novo synthesis of DNA containing the synonymous codons and/or codon pairs and substitution of the corresponding region of the genome with the synthesized DNA. In further embodiments, the entire genome is substituted with the synthesized DNA. In still further embodiments, a portion of the genome is substituted with the synthesized DNA. In yet other embodiments, said portion of the genome is the capsid coding region.

In addition, the present invention provides a method for eliciting a protective immune response in a subject comprising administering to the subject a prophylactically or therapeutically effective dose of any of the vaccine compositions described herein. This invention also provides a method for preventing a subject from becoming afflicted with a virus-associated disease comprising administering to the subject a prophylactically effective dose of any of the instant vaccine compositions. In embodiments of the above methods, the subject has been exposed to a pathogenic virus.

51

"Exposed" to a pathogenic virus means contact with the virus such that infection could result.

The invention further provides a method for delaying the onset, or slowing the rate of progression, of a virus-associated disease in a virus-infected subject comprising administering to the subject a therapeutically effective dose of any of the instant vaccine compositions.

As used herein, "administering" means delivering using any of the various methods and delivery systems known to those skilled in the art. Administering can be performed, for example, intraperitoneally, intracerebrally, intravenously, orally, transmucosally, subcutaneously, transdermally, intra-dermally, intramuscularly, topically, parenterally, via implant, intrathecally, intralymphatically, intralesionally, pericardially, or epidurally. An agent or composition may also be administered in an aerosol, such as for pulmonary and/or intranasal delivery. Administering may be performed, for example, once, a plurality of times, and/or over one or more extended periods.

Eliciting a protective immune response in a subject can be accomplished, for example, by administering a primary dose of a vaccine to a subject, followed after a suitable period of time by one or more subsequent administrations of the vaccine. A suitable period of time between administrations of the vaccine may readily be determined by one skilled in the art, and is usually on the order of several weeks to months. The present invention is not limited, however, to any particular method, route or frequency of administration.

A "subject" means any animal or artificially modified animal. Animals include, but are not limited to, humans, non-human primates, cows, horses, sheep, pigs, dogs, cats, rabbits, ferrets, rodents such as mice, rats and guinea pigs, and birds. Artificially modified animals include, but are not limited to, SCID mice with human immune systems, and CD155tg transgenic mice expressing the human poliovirus receptor CD155. In a preferred embodiment, the subject is a human. Preferred embodiments of birds are domesticated poultry species, including, but not limited to, chickens, turkeys, ducks, and geese.

A "prophylactically effective dose" is any amount of a vaccine that, when administered to a subject prone to viral infection or prone to affliction with a virus-associated disorder, induces in the subject an immune response that protects the subject from becoming infected by the virus or afflicted with the disorder. "Protecting" the subject means either reducing the likelihood of the subject's becoming infected with the virus, or lessening the likelihood of the disorder's onset in the subject, by at least two-fold, preferably at least ten-fold. For example, if a subject has a 1% chance of becoming infected with a virus, a two-fold reduction in the likelihood of the subject becoming infected with the virus would result in the subject having a 0.5% chance of becoming infected with the virus. Most preferably, a "prophylactically effective dose" induces in the subject an immune response that completely prevents the subject from becoming infected by the virus or prevents the onset of the disorder in the subject entirely.

As used herein, a "therapeutically effective dose" is any amount of a vaccine that, when administered to a subject afflicted with a disorder against which the vaccine is effective, induces in the subject an immune response that causes the subject to experience a reduction, remission or regression of the disorder and/or its symptoms. In preferred

52

embodiments, recurrence of the disorder and/or its symptoms is prevented. In other preferred embodiments, the subject is cured of the disorder and/or its symptoms.

Certain embodiments of any of the instant immunization and therapeutic methods further comprise administering to the subject at least one adjuvant. An "adjuvant" shall mean any agent suitable for enhancing the immunogenicity of an antigen and boosting an immune response in a subject. Numerous adjuvants, including particulate adjuvants, suitable for use with both protein- and nucleic acid-based vaccines, and methods of combining adjuvants with antigens, are well known to those skilled in the art. Suitable adjuvants for nucleic acid based vaccines include, but are not limited to, Quil A, imiquimod, resiquimod, and interleukin-12 delivered in purified protein or nucleic acid form. Adjuvants suitable for use with protein immunization include, but are not limited to, alum, Freund's incomplete adjuvant (FIA), saponin, Quil A, and QS-21.

The invention also provides a kit for immunization of a subject with an attenuated virus of the invention. The kit comprises the attenuated virus, a pharmaceutically acceptable carrier, an applicator, and an instructional material for the use thereof. In further embodiments, the attenuated virus may be one or more poliovirus, one or more rhinovirus, one or more influenza virus, etc. More than one virus may be preferred where it is desirable to immunize a host against a number of different isolates of a particular virus. The invention includes other embodiments of kits that are known to those skilled in the art. The instructions can provide any information that is useful for directing the administration of the attenuated viruses.

Of course, it is to be understood and expected that variations in the principles of invention herein disclosed can be made by one skilled in the art and it is intended that such modifications are to be included within the scope of the present invention. The following Examples further illustrate the invention, but should not be construed to limit the scope of the invention in any way. Detailed descriptions of conventional methods, such as those employed in the construction of recombinant plasmids, transfection of host cells with viral constructs, polymerase chain reaction (PCR), and immunological techniques can be obtained from numerous publications, including Sambrook et al. (1989) and Coligan et al. (1994). All references mentioned herein are incorporated in their entirety by reference into this application.

Full details for the various publications cited throughout this application are provided at the end of the specification immediately preceding the claims. The disclosures of these publications are hereby incorporated in their entireties by reference into this application. However, the citation of a reference herein should not be construed as an acknowledgement that such reference is prior art to the present invention.

EXAMPLE 1

Re-Engineering of Capsid Region of Polioviruses by
60 Altering Codon Bias

Cells, Viruses, Plasmids, and Bacteria
HeLa R19 cell monolayers were maintained in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% bovine calf serum (BCS) at 37° C. All PV infectious cDNA constructs are based on PV1(M) cDNA clone pT7PVM (Cao et al., 1993; van der Werf et al., 1986). Dicistronic reporter plasmids were constructed using

pHRPF-Luc (Zhao and Wimmer, 2001). *Escherichia coli* DH5 α was used for plasmid transformation and propagation. Viruses were amplified by infection of HeLa R19 cell monolayers with 5 PFU per cell. Infected cells were incubated in DMEM (2% BCS) at 37° C. until complete cytopathic effect (CPE) was apparent or for at least 4 days post-infection. After three rounds of freezing and thawing, the lysate was clarified of cell debris by low-speed centrifugation and the supernatant, containing the virus, was used for further passaging or analysis.

Cloning of Synthetic Capsid Replacements and Dicistronic Reporter Replicons

Two PV genome cDNA fragments spanning the genome between nucleotides 495 and 3636, named SD and AB, were synthesized using GeneMaker® technology (Blue Heron Biotechnology). pPV-SD and pPV-AB were generated by releasing the replacement cassettes from the vendor's cloning vector by PflMI digestion and insertion into the pT7PVM vector in which the corresponding PflMI fragment had been removed. pPV-AB⁷⁵⁵⁻¹⁵¹³ and pPV-AB²⁴⁷⁰⁻³³⁸⁶ were obtained by inserting a BsmI fragment or an NheI-EcoRI fragment, respectively, from pPV-AB into equally digested pT7PVM vector. In pPV-AB¹⁵¹³⁻³³⁸⁶ and pPV-AB⁷⁵⁵⁻²⁴⁷⁰, the BsmI fragment or NheI-EcoRI fragment of pT7PVM, respectively, replaces the respective fragment of the pPV-AB vector. Replacement of the NheI-EcoRI fragment of pPV-AB¹⁵¹³⁻³³⁸⁶ with that of pT7PVM resulted in pPV-AB²⁴⁷⁰⁻³³⁸⁶. Finally, replacement of the SnaBI-EcoRI fragments of pPV-AB²⁴⁷⁰⁻³³⁸⁶ and pT7PVM with one another produced pPV-AB²⁹⁵⁴⁻³³⁸⁶ and pPV-AB²⁴⁷⁰⁻²⁹⁵⁴, respectively.

Cloning of dicistronic reporter constructs was accomplished by first introducing a silent mutation in pHRPF-Luc by site-directed mutagenesis using oligonucleotides Fluc-mutRI(+) / Fluc-mutRI(-) to mutate an EcoRI site in the firefly luciferase open reading frame and generate pdiLuc-mRI. The capsid regions of pT7PVM, pPV-AB¹⁵¹³⁻²⁴⁷⁰, and pPV-AB²⁴⁷⁰⁻²⁹⁵⁴ were PCR amplified using oligonucleotides RI-2A-P1wt(+) / P1wt-2A-RI(-). Capsid sequences of pPV-AB²⁴⁷⁰⁻³³⁸⁶ and pPV-AB²⁹⁵⁴⁻³³⁸⁶ or pPV-AB were amplified with RI-2A-P1wt(+) / P1AB-2A-RI(-) or RI-2A-P1AB(+) / P1AB-2A-RI(-), respectively. PCR products were digested with EcoRI and inserted into a now unique EcoRI site in pdiLuc-mRI to result in pdiLuc-PV, pdiLuc-AB¹⁵¹³⁻²⁴⁷⁰, pdiLuc-AB²⁴⁷⁰⁻²⁹⁵⁴, pdiLuc-AB²⁴⁷⁰⁻³³⁸⁶, pdiLuc-AB²⁹⁵⁴⁻³³⁸⁶, and pdiLuc-AB, respectively.

Oligonucleotides

The following oligonucleotides were used:

(SEQ ID NO: 6)

Fluc-mutRI(+),
5'-GCACTGATAATGAACCTCCTCTGGATCTACTGG-3';

(SEQ ID NO: 7)

Fluc-mutRI(-),
5'-CCAGTAGATCCAGAGGAGTCATTATCAGTGC-3';

(SEQ ID NO: 8)

RI-2A-P1wt(+),
5'-CAAGAATTCTGACCACATACGGTGCTCAGGTTCATCACAGAAAGT
GGG-3';

(SEQ ID NO: 9)

RI-2A-P1AB(+),
5'-CAAGAATTCTGACCACATACGGTGCGCAAGTATCGTCGCAAAAAGT
AGG-3';

-continued

(SEQ ID NO: 10)

P1wt-2A-RI(-),
5'-TCGAATTCTCCATATGTGGTCAGATCCTTGGTGG-AGAGG-3';
and

5

(SEQ ID NO: 11)

P1AB-2A-RI(-),
5'-TCGAATTCTCCATACGTCGTTAAATCTTCGTCGATAACG-3'.

In vitro Transcription and RNA Transfection

10 Driven by the T7 promoter, 2 µg of EcoRI-linearized plasmid DNA were transcribed by T7 RNA polymerase (Stratagene) for 1 h at 37° C. One microgram of virus or dicistronic transcript RNA was used to transfect 10⁶ HeLa R19 cells on a 35-mm-diameter plate according to a modification of the DEAE-dextran method (van der Werf et al., 1986). Following a 30-min incubation at room temperature, the supernatant was removed and cells were incubated at 37° C. in 2 ml of DMEM containing 2% BCS until CPE appeared, or the cells were frozen 4 days post-transfection for further passaging. Virus titers were determined by standard plaque assay on HeLa R19 cells using a semisolid overlay of 0.6% agar (Sigma-Aldrich) in minimal Eagle medium.

15 20 25 Design and Synthesis of Codon-Deoptimized Polioviruses

Two different synonymous encodings of the poliovirus P1 capsid region were produced, each governed by different design criteria. The designs were limited to the capsid, as it has been conclusively shown that the entire capsid coding sequence can be deleted from the PV genome or replaced with exogenous sequences without affecting replication of the resulting sub-genomic replicon (Johansen and Morrow, 2000; Kaplan and Racaniello, 1988). It is therefore quite certain that no unidentified crucial regulatory RNA elements are located in the capsid region, which might be affected inadvertently by modulation of the RNA sequence.

30 35 40 45 50 55 60 65 70 75 80 85 90 95 The first design (PV-SD) sought to maximize the number of RNA base changes while preserving the exact codon usage distribution of the wild type P1 region (FIG. 1). To achieve this, synonymous codon positions were exchanged for each amino acid by finding a maximum weight bipartite match (Gabow, 1973) between the positions and the codons, where the weight of each position-codon pair is the number of base changes between the original codon and the synonymous candidate codon to replace it. To avoid any positional bias from the matching algorithm, the synonymous codon locations were randomly permuted before creating the input graph and the locations were subsequently restored. Rothberg's maximum bipartite matching program (Rothberg, 1985) was used to compute the matching. A total of 11 useful restriction enzyme sites, each 6 nucleotides, were locked in the viral genome sequence so as to not participate in the codon location exchange. The codon shuffling technique potentially creates additional restriction sites that should preferably remain unique in the resulting reconstituted full-length genome. For this reason, the sequence was further processed by substituting codons to eliminate the undesired sites. This resulted in an additional nine synonymous codon changes that slightly altered the codon frequency distribution. However, no codon had its frequency changed by more than 1 over the wild type sequence. In total, there were 934 out of 2,643 nucleotides changed in the 60 65 70 75 80 85 90 95 PV-SD capsid design when compared to the wt P1 sequence while maintaining the identical protein sequence of the capsid coding domain (see FIGS. 1 and 2). As the codon

usage was not changed, the GC content in the PVM-SD capsid coding sequence remained identical to that in the wt at 49%.

The second design, PV-AB, sought to drastically change the codon usage distribution over the wt P1 region. This design was influenced by recent work suggesting that codon bias may impact tissue-specific expression (Plotkin et al., 2004). The desired codon usage distribution was derived from the most unfavorable codons observed in a previously described set of brain-specific genes (Hsiao et al., 2001; Plotkin et al., 2004). A capsid coding region was synthesized maximizing the usage of the rarest synonymous codon for each particular amino acid as observed in this set of genes (FIG. 1). Since for all amino acids but one (Leu) the rarest codon in brain corresponds to the rarest codons among all human genes at large, in effect this design would be expected to discriminate against expression in other human tissues as well. Altogether, the PV-AB capsid differs from the wt capsid in 680 nucleotide positions (see FIG. 2). The GC content in the PVM-AB capsid region was reduced to 43% compared to 49% in the wt.

EXAMPLE 2

Effects of Codon-Deoptimization on Growth and Infectivity of Polioviruses

Determination of Virus Titer by Infected Focus Assay

Infections were done as for a standard plaque assay. After 48 or 72 h of incubation, the tragacanth gum overlay was removed and the wells were washed twice with phosphate-buffered saline (PBS) and fixed with cold methanol/acetone for 30 min. Wells were blocked in PBS containing 10% BCS followed by incubation with a 1:20 dilution of anti-3D mouse monoclonal antibody 125.2.3 (Paul et al., 1998) for 1 h at 37° C. After washing, cells were incubated with horseradish peroxidase-labeled goat anti-mouse antibody (Jackson ImmunoResearch, West Grove, Pa.) and infected cells were visualized using Vector VIP substrate kit (Vector Laboratories, Burlingame, Calif.). Stained foci, which are equivalent to plaques obtained with wt virus, were counted, and titers were calculated as in the plaque assay procedure.

Codon-Deoptimized Polioviruses Display Severe Growth Phenotypes

Of the two initial capsid ORF replacement designs (FIG. 3A), only PV-SD produced viable virus. In contrast, no viable virus was recovered from four independent transfections with PV-AB RNA, even after three rounds of passaging (FIG. 3E). It appeared that the codon bias introduced into the PV-AB genome was too severe. Thus, smaller portions of the PV-AB capsid coding sequence were subcloned into the PV(M) background to reduce the detrimental effects of the nonpreferred codons. Of these subclones, PV-AB²⁹⁵⁴⁻³³⁸⁶ produced CPE 40 h after RNA transfaction, while PV-AB⁷⁵⁵⁻¹⁵¹³ and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ required one or two additional passages following transfection, respectively (compared to 24 h for the wild type virus). Interestingly, these chimeric viruses represent the three subclones with the smallest portions of the original AB sequence, an observation suggesting a direct correlation between the number of nonpreferred codons and the fitness of the virus.

One-step growth kinetics of all viable virus variants were determined by infecting HeLa monolayers at a multiplicity of infection (MOI) of 2 with viral cell lysates obtained after a maximum of two passages following RNA transfection

(FIG. 3B). The MOI was chosen due to the low titer of PV-AB²⁴⁷⁰⁻²⁹⁵⁴ and to eliminate the need for further passaging required for concentrating and purifying the inoculum. Under the conditions used, all viruses had produced complete or near complete CPE by 24 h post-infection.

Despite 934 single-point mutations in its capsid region, PV-SD replicated at wt capacity (FIG. 3B) and produced similarly sized plaques as the wt (FIG. 3D). While PV-AB²⁹⁵⁴⁻³³⁸⁶ grew with near-wild type kinetics (FIG. 3B), PV-AB⁷⁵⁵⁻¹⁵¹³ produced minute plaques and approximately 22-fold less infectious virus (FIGS. 2, 3B and F, respectively). Although able to cause CPE in high-MOI infections, albeit much delayed (80 to 90% CPE after 20 to 24 h), PV-AB²⁴⁷⁰⁻²⁹⁵⁴ produced no plaques at all under the conditions of the standard plaque assay (FIG. 3H). This virus was therefore quantified using a focus-forming assay, in which foci of infected cells after 72 h of incubation under plaque assay conditions were counted after they were stained immunohistochemically with antibodies to the viral polymerase 3D (FIG. 3G). After 48 h of infection, PV-AB²⁴⁷⁰⁻²⁹⁵⁴-infected foci usually involved only tens to hundreds of cells (FIG. 3J) with a focus diameter of 0.2 to 0.5 mm, compared to 3-mm plaques for the wt (FIGS. 3C and D). However, after an additional 24 h, the diameter of the foci increased significantly (2 to 3 mm; FIG. 3G). When HeLa cells were infected with PV-AB⁷⁵⁵⁻¹⁵¹³ and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ at an MOI of 1, the CPE appeared between 12 and 18 h and 3 and 4 days, respectively, compared to 8 h with the wt (data not shown).

In order to quantify the cumulative effect of a particular codon bias in a protein coding sequence, a relative codon deoptimization index (RCDI) was calculated, which is a comparative measure against the general codon distribution in the human genome. An RCDI of 1/codon indicates that a gene follows the normal human codon frequencies, while any deviation from the normal human codon bias results in an RCDI higher than 1. The RCDI was derived using the formula:

$$\text{RCDI} = \left[\sum (C_i F_a / C_i F_h) N_{ci} \right] / N \quad (i=1 \text{ through } 64)$$

$C_i F_a$ is the observed relative frequency in the test sequence of each codon i out of all synonymous codons for the same amino acid (0 to 1); $C_i F_h$ is the normal relative frequency observed in the human genome of each codon i out of all synonymous codons for that amino acid (0.06 to 1); N_{ci} is the number of occurrences of that codon i in the sequence; and N is the total number of codons (amino acids) in the sequence.

Thus, a high number of rare codons in a sequence results in a higher index. Using this formula, the RCDI values of the various capsid coding sequences were calculated to be 1.14 for PV(M) and PV-SD which is very close to a normal human distribution. The RCDI values for the AB constructs are 1.73 for PV-AB⁷⁵⁵⁻¹⁵¹³, 1.45 for PV-AB²⁴⁷⁰⁻²⁹⁵⁴, and 6.51 for the parental PV-AB. For comparison, the RCDI for probably the best known codon-optimized protein, "humanized" green fluorescent protein (GFP), was 1.31 compared to an RCDI of 1.68 for the original *Aequorea victoria* gfp gene (Zolotukhin et al., 1996). According to these calculations, a capsid coding sequence with an RCDI of <2 is associated with a viable virus phenotype, while an RCDI of >2 (PV-AB=6.51, PV-AB¹⁵¹³⁻³³⁸⁶=4.04, PV-AB⁷⁵⁵⁻²⁴⁷⁰=3.61) results in a lethal phenotype.

Effects of Codon-Deoptimization on Specific Infectivity of Polioviruses

Molecular Quantification of Viral Particles: Direct OD₂₆₀ Absorbance Method

Fifteen-centimeter dishes of HeLa cells (4×10^7 cells) were infected with PV(M), PV-AB⁷⁵⁵⁻¹⁵¹³, or PV-AB²⁴⁷⁰⁻²⁹⁵⁴ at an MOI of 0.5 until complete CPE occurred (overnight versus 4 days). Cell-associated virus was released by three successive freeze/thaw cycles. Cell lysates were cleared by 10 min of centrifugation at 2,000×g followed by a second 10-min centrifugation at 14,000×g for 10 min. Supernatants were incubated for 1 h at room temperature in the presence of 10 µg/ml RNase A (Roche) to digest any extraviral or cellular RNA. After addition of 0.5% sodium dodecyl sulfate (SDS) and 2 mM EDTA, virus-containing supernatants were overlaid on a 6-ml sucrose cushion (30% sucrose in Hanks balanced salt solution [HBSS]; Invitrogen, Carlsbad, Calif.). Virus particles were sedimented by ultracentrifugation for 4 h at 28,000 rpm using an SW28 swinging bucket rotor. Supernatants were discarded and centrifuge tubes were rinsed twice with HBSS while leaving the sucrose cushion intact. After removal of the last wash and the sucrose cushion, virus pellets were resuspended in PBS containing 0.2% SDS and 5 mM EDTA. Virus infectious titers were determined by plaque assay/infected-focus assay (see above). Virus particle concentrations were determined with a NanoDrop spectrophotometer (NanoDrop Technologies, Inc., Wilmington, Del.) at the optical density at 260 nm (OD₂₆₀) and calculated using the formula 1 OD₂₆₀ unit=9.4×10¹² particles/ml (Rueckert, 1985). In addition, virion RNA was extracted by three rounds of phenol extraction and one round of chloroform extraction. RNA was ethanol precipitated and resuspended in ultrapure water. RNA purity was confirmed by TAE-buffered agarose gel analysis, and the concentration was determined spectrophotometrically. The total number of genome equivalents of the corresponding virus preparation was calculated via the determined RNA concentration and the molecular weight of the RNA. Thus, the relative amount of virions per infectious units could be calculated, assuming that one RNase-protected viral genome equivalent corresponds to one virus particle.

Molecular Quantification of Viral Particles: ELISA Method

Nunc Maxisorb 96-well plates were coated with 10 µg of rabbit anti-PV(M) antibody (Murdin and Wimmer, 1989) in 100 µl PBS for 2 h at 37° C. and an additional 14 h at 4° C., and then the plates were washed three times briefly with 350 µl of PBS and blocked with 350 µl of 10% bovine calf serum in PBS for 1 h at 37° C. Following three brief washes with PBS, wells were incubated with 100 µl of virus-containing cell lysates or controls in DMEM plus 2% BCS for 4 h at room temperature. Wells were washed with 350 µl of PBS three times for 5 min each. Wells were then incubated for 4 h at room temperature with 2 µg of CD155-alkaline phosphatase (AP) fusion protein (He et al., 2000) in 100 µl of DMEM-10% BCS. After the last of five washes with PBS, 100 µl of 10 mM Tris, pH 7.5, were added and plates were incubated for 1 h at 65° C. Colorimetric alkaline phosphatase determination was accomplished by addition of 100 µl of 9 mg/ml para-nitrophenylphosphate (in 2 M diethanolamine, 1 mM MgCl₂, pH 9.8). Alkaline phosphatase activity was determined, and virus particle concentrations were calculated in an enzyme-linked immunosorbent assay (ELISA) plate reader (Molecular Devices, Sunnyvale, Calif.) at a 405-nm wavelength on a standard curve prepared

in parallel using two-fold serial dilutions of a known concentration of purified PV(M) virus stock.

The PFU/Particle Ratio is Reduced in Codon-Deoptimized Viruses

The extremely poor growth phenotype of PV-AB²⁴⁷⁰⁻²⁹⁵⁴ in cell culture and its inability to form plaques suggested a defect in cell-to-cell spreading that may be consistent with a lower specific infectivity of the individual virus particles. To test this hypothesis, PV(M), PV-AB⁷⁵⁵⁻¹⁵¹³, and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ virus were purified and the amount of virus particles was determined spectrophotometrically. Purified virus preparations were quantified directly by measuring the OD₂₆₀, and particle concentrations were calculated according to the formula 1 OD₂₆₀ unit=9.4×10¹² particles/ml (FIG. 4D) (Rueckert, 1985). Additionally, genomic RNA was extracted from those virions (FIG. 4A) and quantified at OD₂₆₀ (data not shown). The number of virions (1 virion=1 genome) was then determined via the molecular size of 2.53×10⁶ g/mol for genomic RNA. Specifically, virus was prepared from 4×10^7 HeLa cells that were infected with 0.5 MOI of virus until the appearance of complete CPE, as described above. Both methods of particle determinations produced similar results (FIG. 4D). Indeed, it was found that PV(M) and PV-AB⁷⁵⁵⁻¹⁵¹³ produced roughly equal amounts of virions, while PV-AB²⁴⁷⁰⁻²⁹⁵⁴ produced between $\frac{1}{3}$ (by the direct UV method (FIG. 4D) to $\frac{1}{6}$ of the number of virions compared to PV(M) (by genomic RNA method [data not shown]). In contrast, the wt virus sample corresponded to approximately 30 times and 3,000 times more infectious units than PV-AB⁷⁵⁵⁻¹⁵¹³ and PV-AB²⁴⁷⁰⁻²⁹⁵⁴, respectively (FIG. 4D). In addition, capsid proteins of purified virions were resolved by SDS-polyacrylamide gel electrophoresis (PAGE) and visualized by silver staining (FIG. 4B). These data also support the conclusion that on a per-cell basis, PV-AB²⁴⁷⁰⁻²⁹⁵⁴ and PV-AB⁷⁵⁵⁻¹⁵¹³ produce similar or only slightly reduced amounts of progeny per cell (FIG. 4B, lane 3), while their PFU/particle ratio is reduced. The PFU/particle ratio for a virus can vary significantly depending on the methods to determine either plaques (cell type for plaque assay and the particular plaque assay technique) or particle count (spectrophotometry or electron microscopy). A PFU/particle ratio of 1/115 for PV1(M) was determined using the method described herein, which compares well to previous determinations of 1/272 (Joklik and Darnell, 1961) (done on HeLa cells) and 1/87 (Schwerdt and Fogh, 1957) (in primary monkey kidney cells).

Development of a Virion-Specific ELISA

To confirm the reduced PFU/particle ratio observed with codon-deoptimized polioviruses, a novel virion-specific ELISA was developed (FIGS. 4C and E) as a way to determine the physical amount of intact viral particles in a sample rather than the infectious titer, which is a biological variable. The assay is based on a previous observation that the ectodomain of the PV receptor CD155 fused to heat-stable placental alkaline phosphatase (CD155-AP) binds very tightly and specifically to the intact 160S particle (He et al., 2000). Considering that PV 130S particles (A particles) lose their ability to bind CD155 efficiently (Hogle, 2002), it is expected that no other capsid intermediate or capsid subunits would interact with CD155-AP, thus ensuring specificity for intact particles. In support of this notion, lysates from cells that were infected with a vaccinia virus strain expressing the P1 capsid precursor (Ansardi et al., 1993) resulted in no quantifiable signal (data not shown).

The ELISA method allows for the quantification of virus particles in a crude sample such as the cell lysate after infection, which should minimize possible alteration of the

59

PFU/particle ratio by other mechanisms during sample handling and purification (thermal/chemical inactivation, oxidation, degradation, etc.). Under the current conditions, the sensitivity of this assay is approximately 10^7 viral particles, as there is no signal amplification step involved. This, in turn, resulted in an exceptionally low background. With this ELISA, PV particle concentrations could be determined in samples by back calculation on a standard curve prepared with purified PV(M) of known concentration (FIG. 4E). The particle determinations by ELISA agreed well with results obtained by the direct UV method (FIG. 4D).

Implications of Results

The present study has demonstrated the utility of large-scale codon deoptimization of PV capsid coding sequences by de novo gene synthesis for the generation of attenuated viruses. The initial goal was to explore the potential of this technology as a tool for generating live attenuated virus vaccines. Codon-deoptimized viruses were found to have

60

intracerebrally with virus dilutions between 10^2 and 10^6 PFU/focus-forming units (FFU) in 30 μ l PBS. Fifty percent lethal dose (LD_{50}) values were calculated by the method of Reed and Muench (1938). Virus titers in spinal cord tissues at the time of death or paralysis were determined by plaque or infected-focus assay.

Codon-Deoptimized Polioviruses are Neuroattenuated on a Particle Basis in CD155tg Mice

To test the pathogenic potential of viruses constructed in this study, CD155 transgenic mice (Koike et al., 1991) were injected intracerebrally with PV(M), PV-SD, PV-AB⁷⁵⁵⁻¹⁵¹³, and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ at doses between 10^2 and 10^5 PFU/FFU. Initial results were perplexing, as quite counter-intuitively PV-AB⁷⁵⁵⁻¹⁵¹³ and especially PV-AB²⁴⁷⁰⁻²⁹⁵⁴ were initially found to be as neuropathogenic as, or even slightly more neuropathogenic, than the wt virus. See Table 4.

TABLE 4

Construct	Neuropathogenicity in CD155tg mice.			
	LD ₅₀	Spinal cord titer		
	PFU or FFU ^a	No. of virions ^b	PFU or FFU/g ^c	No. of virions/g ^d
PV(M) wt	3.2×10^2 PFU	3.7×10^4	1.0×10^9 PFU	1.15×10^{11}
PV-AB ⁷⁵⁵⁻¹⁵¹³	2.6×10^2 PFU	7.3×10^5	3.5×10^7 PFU	9.8×10^{10}
PV-AB ²⁴⁷⁰⁻²⁹⁵⁴	4.6×10^2 PFU	4.8×10^6	3.4×10^6 FFU	3.57×10^{11}

^aLD₅₀ expressed as the number of infectious units, as determined by plaque or infectious focus assay, that results in 50% lethality after intracerebral inoculation.

^bLD₅₀ expressed as the number of virus particles, as determined by OD₂₆₀ measurement, that results in 50% lethality after intracerebral inoculation.

^cVirus recovered from the spinal cord of infected mice at the time of death or paralysis; expressed in PFU or FFU/g of tissue, as determined by plaque or infectious focus assay.

^dVirus recovered from the spinal cord of infected mice at the time of death or paralysis, expressed in particles/g of tissue, derived by multiplying values in the third column by the particle/PFU ratio characteristic for each virus (FIG. 4D).

35

very low specific infectivity (FIG. 4). The low specific infectivity (that is the chance of a single virus particle to successfully initiate an infectious cycle in a cell) results in a more slowly spreading virus infection within the host. This in turn allows the host organism more time to mount an immune response and clear the infection, which is a most desirable feature in an attenuated virus vaccine. On the other hand, codon-deoptimized viruses generated similar amounts of progeny per cell as compared the wild type virus, while being 2 to 3 orders of magnitude less infectious (FIG. 4). This allows the production of virus particles antigenically indistinguishable from the wt as effectively and cost-efficiently as the production of the wt virus itself. However due to the low specific infectivity the actual handling and processing of such a virus preparation is much safer. Since, there are increasing concerns about the production of virulent virus in sufficient quantities under high containment conditions and the associated risk of virus escape from the production facility either by accident or by malicious intent. Viruses as described herein may prove very useful as safer alternatives in the production of inactivated virus vaccines. Since they are 100% identical to the wt virus at the protein level, an identical immune response in hosts who received inactivated virus is guaranteed.

EXAMPLE 4

Effects of Codon-Deoptimization on Neuropathogenicity of Polioviruses

Mouse Neuropathogenicity Tests

Groups of four to five CD155tg mice (strain Tg21) (Koike et al., 1991) between 6 and 8 weeks of age were injected

In addition, times of onset of paralysis following infection with PV-AB⁷⁵⁵⁻¹⁵¹³ and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ were comparable to that of wt virus (data not shown). Similarly confounding was the observation that at the time of death or paralysis, the viral loads, as determined by plaque assay, in the spinal cords of mice infected with PV-AB⁷⁵⁵⁻¹⁵¹³ and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ were 30- and 300-fold lower, respectively, than those in the mice infected with the wt virus (Table 4). Thus, it seemed unlikely that PV-AB²⁴⁷⁰⁻²⁹⁵⁴, apparently replicating at only 0.3% of the wt virus, would have the same neuropathogenic potential as the wt. However, after having established the altered PFU/particle relationship in PV-AB⁷⁵⁵⁻¹⁵¹³ and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ (see Example 3), the amount of inoculum could now be correlated with the actual number of particles inoculated. After performing this correction, it was established that on a particle basis, PV-AB⁷⁵⁵⁻¹⁵¹³ and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ are 20-fold and 100-fold neuroattenuated, respectively, compared to the wt. See Table 4. Furthermore, on a particle basis the viral loads in the spinal cords of paralyzed mice were very similar with all three viruses (Table 4).

It was also concluded that it was not possible to redesign the PV capsid gene with synonymous codons that would specifically discriminated against expression in the central nervous system. This may be because tissue-specific differences in codon bias described by others (Plotkin et al., 2004) are too small to bring about a tissue-restrictive virus phenotype. In a larger set of brain-specific genes than the one used by Plotkin et al., no appreciable tissue-specific codon bias was detected (data not shown). However, this conclusion should not detract from the fact that polioviruses produced by the method of this invention are indeed neu-

61

roattenuated in mice by a factor of up to 100 fold. That is, 100 fold more of the codon or codon-pair deoptimized viral particles are needed to result in the same damage in the central nervous system as the wt virus.

EXAMPLE 5

Effects of Codon Deoptimization on Genomic Translation of Polioviruses

In Vitro and In Vivo Translation

Two different HeLa cell S10 cytoplasmic extracts were used in this study. A standard extract was prepared by the method of Molla et al. (1991). [³⁵S]methionine-labeled translation products were analyzed by gel autoradiography. The second extract was prepared as described previously (Kaplan and Racaniello, 1988), except that it was not dialyzed and endogenous cellular mRNAs were not removed with micrococcal nuclease. Reactions with the modified extract were not supplemented with exogenous amino acids or tRNAs. Translation products were analyzed by western blotting with anti-2C monoclonal antibody 91.23 (Pfister and Wimmer, 1999). Relative intensities of 2BC bands were determined by a pixel count of the scanned gel image using the NIH-Image 1.62 software. In all cases, translation reactions were programmed with 200 ng of the various in vitro-transcribed viral genomic RNAs.

For analysis of in vivo translation, HeLa cells were transfected with in vitro-transcribed dicistronic replicon RNA as described above. In order to assess translation isolated from RNA replication, transfections were carried out in the presence of 2 mM guanidine hydrochloride. Cells were lysed after 7 h in passive lysis buffer (Promega, Madison, Wis.) followed by a dual firefly (F-Luc) and *Renilla* (R-Luc) luciferase assay (Promega). Translation efficiency of the second cistron (P1-Fluc-P2-P3 polyprotein) was normalized through division by the *Renilla* luciferase activity of the first cistron expressed under control of the Hepatitis C Virus (HCV) internal ribosome entry site (IRES).

Codon-Deoptimized Viruses are Deficient at the Level of Genome Translation

Since the synthetic viruses and the wt PV(M) are indistinguishable in their protein makeup and no known RNA-based regulatory elements were altered in the modified RNA genomes, these designs enabled study of the effect of reduced genome translation/replication on attenuation without affecting cell and tissue tropism or immunological properties of the virus. The PV-AB genome was designed under the hypothesis that introduction of many suboptimal codons into the capsid coding sequence should lead to a reduction of genome translation. Since the P1 region is at the N-terminus of the polyprotein, synthesis of all downstream nonstructural proteins is determined by the rate of translation through the P1 region. To test whether in fact translation is affected, in vitro translations were performed (FIG. 5).

Unexpectedly, the initial translations in a standard HeLa-cell based cytoplasmic S10 extract (Molla et al., 1991) showed no difference in translation capacities for any of the genomes tested (FIG. 5A). However, as this translation system is optimized for maximal translation, it includes the exogenous addition of excess amino acids and tRNAs, which could conceivably compensate for the genetically engineered codon bias. Therefore, in vitro translations were repeated with a modified HeLa cell extract, which was not dialyzed and in which cellular mRNAs were not removed by micrococcal nuclease treatment (FIG. 5B). Translations in this extract were performed without the addition of exog-

62

enous tRNAs or amino acids. Thus, an environment was created that more closely resembles that in the infected cell, where translation of the PV genomes relies only on cellular supplies while competing for resources with cellular mRNAs. Due to the high background translation from cellular mRNA and the low [³⁵S]Met incorporation rate in nondialyzed extract, a set of virus-specific translation products were detected by western blotting with anti-2C antibodies (Pfister and Wimmer, 1999). These modified conditions resulted in dramatic reduction of translation efficiencies of the modified genomes which correlated with the extent of the deoptimized sequence. Whereas translation of PV-SD was comparable to that of the wt, translation of three noninfectious genomes, PV-AB, PV-AB¹⁵¹³⁻³³⁸⁶, and PV-AB⁷⁵⁵⁻²⁴⁷⁰, was reduced by approximately 90% (FIG. 5B).

Burns et al. (2006) recently reported experiments related to those described herein. These authors altered codon usage to a much more limited extent than in the present study, and none of their mutant viruses expressed a lethal phenotype. Interestingly, Burns et al. determined that translation did not play a major role in the altered phenotypes of their mutant viruses, a conclusion at variance with the data presented herein. It is likely that the in vitro translation assay used by Burns et al. (2006), which employed a nuclease-treated rabbit reticulocyte lysate supplemented with uninfected HeLa cell extract and excess amino acids, explains their failure to detect any significant reduction in translation. Cf FIG. 5A.

Considering the ultimately artificial nature of the in vitro translation system, the effect of various capsid designs on translation in cells was also investigated. For this purpose, dicistronic poliovirus reporter replicons were constructed (FIG. 6A) based on a previously reported dicistronic replicon (Zhao and Wimmer, 2001). Various P1 cassettes were inserted immediately upstream and in-frame with the firefly luciferase (F-Luc) gene. Thus, the poliovirus IRES drives expression of a single viral polyprotein similar to the one in the viral genome, with the exception of the firefly luciferase protein between the capsid and the 2A^{Pro} proteinase. Expression of the *Renilla* luciferase (R-Luc) gene under the control of the HCV IRES provides an internal control. All experiments were carried out in the presence of 2 mM guanidine hydrochloride, which completely blocks genome replication (Wimmer et al., 1993). Using this type of construct allowed an accurate determination of the relative expression of the second cistron by calculating the F-Luc/R-Luc ratio. As F-Luc expression depends on successful transit of the ribosome through the upstream P1 region, it provides a measure of the effect of the inserted P1 sequence on the rate of polyprotein translation. Using this method, it was indeed found that the modified capsid coding regions, which were associated with a lethal phenotype in the virus background (e.g., PV-AB, PV-AB¹⁵¹³⁻²⁴⁷⁰, and PV-AB²⁴⁷⁰⁻³³⁸⁶) reduced the rate of translation by approximately 80 to 90% (FIG. 6B). Capsids from two viable virus constructs, PV-AB²⁴⁷⁰⁻²⁹⁵⁴ and PV-AB²⁹⁵⁴⁻³³⁸⁶, allowed translation at 68% and 83% of wt levels, respectively. In vivo translation rates of the first cistron remained constant in all constructs over a time period between 3 and 12 h, suggesting that RNA stability is not affected by the codon alterations (data not shown). In conclusion, the results of these experiments suggest that poliovirus is extremely dependent on very efficient translation as a relatively small drop in translation efficiency through the P1 region of 30%, as seen in PV-AB²⁴⁷⁰⁻²⁹⁵⁴, resulted in a severe virus replication phenotype.

63

EXAMPLE 6

Genetic Stability of Codon-Deoptimized Polioviruses

Due to the distributed effect of many mutations over large genome segments that contribute to the phenotype, codon deoptimized viruses should have genetically stable phenotypes. To study the genetic stability of codon deoptimized viruses, and to test the premise that these viruses are genetically stable, viruses are passaged in suitable host cells. A benefit of the present “death by 1000 cuts” theory of vaccine design is the reduced risk of reversion to wild type. Typical vaccine strains differ by only few point mutations from the wt viruses, and only a small subset of these may actually contribute to attenuation. Viral evolution quickly works to revert such a small number of active mutations. Indeed, such reversion poses a serious threat for the World Health Organization (WHO) project to eradicate poliovirus from the globe. So long as a live vaccine strain is used, there is a very real chance that this strain will revert to wt. Such reversion has already been observed as the source of new polio outbreaks (Georgescu et al., 1997; Kew et al., 2002; Shimizu et al., 2004).

With hundreds to thousands of point mutations in the present synthetic designs, there is little risk of reversion to wt strains. However, natural selection is powerful, and upon passaging, the synthetic viruses inevitably evolve. Studies are ongoing to determine the end-point of this evolution, but a likely outcome is that they get trapped in a local optimum, not far from the original design.

To validate this theory, representative re-engineered viruses are passaged in a host cell up to 50 times. The genomes of evolved viruses are sequenced after 10, 20 and 50 passages. More specifically, at least one example chimera from each type of deoptimized virus is chosen. The starting chimera is very debilitated, but not dead. For example, for PV the chimeras could be PV-AB²⁴⁷⁰⁻²⁹⁵⁴ and PV-Min⁷⁵⁵⁻²⁴⁷. From each starting virus ten plaques are chosen. Each of the ten plaque-derived virus populations are bulk passaged a total of 50 times. After the 10th, 20th and 50th passages, ten plaque-purified viruses are again chosen and their genomes are sequenced together with the genomes of the ten parent viruses. After passaging, the fitness of the 40 (30+10 per parent virus) chosen viruses is compared to that of their parents by examining plaque size, and determining plaque forming units/ml as one-step growth kinetics. Select passage isolates are tested for their pathogenicity in appropriate host organisms. For example, the pathogenicity of polioviruses is tested in CD155tg mice.

Upon sequencing of the genomes, a finding that all 10 viral lines have certain mutations in common would suggest that these changes are particularly important for viral fitness. These changes may be compared to the sites identified by toeprinting as the major pause sites (see Example 9); the combination of both kinds of assay may identify mutant codons that are most detrimental to viral fitness. Conversely, a finding that the different lines have all different mutations would support the view that many of the mutant codon changes are very similar in their effect on fitness. Thus far, after 10 passages in HeLa cells, PV-AB⁷⁵⁵⁻¹⁵¹³ and PV-AB²⁴⁷⁰⁻²⁹⁵⁴ have not undergone any perceivable gain of fitness. Viral infectious titers remained as low (10⁷ PFU/ml and 10⁶ FFU/ml) as at the beginning of the passage experiment, and plaque phenotype did not change (data not shown). Sequence analysis of these passaged viruses is now in progress, to determine if and what kind of genetic changes occur during passaging.

64

Burns et al. (2006) reported that their altered codon compositions were largely conserved during 25 serial passages in HeLa cells. They found that whereas the fitness for replication in HeLa cells of both the unmodified Sabin 2 virus and the codon replacement viruses increased with higher passage numbers, the relative fitness of the modified viruses remained lower than that of the unmodified virus. Thus, all indications are that viruses redesigned by SAVE are genetically very stable. Preliminary data for codon and codon-pair deoptimized viruses of the invention suggest that less severe codon changes distributed over a larger number of codons improves the genetic stability of the individual virus phenotypes and thus improves their potential for use in vaccines.

EXAMPLE 7

Re-Engineering of Capsid Region of Polioviruses by Deoptimizing Codon Pairs

Calculation of Codon Pair Bias.

Every individual codon pair of the possible 3721 non-“STOP” containing codon pairs (e.g., GTT-GCT) carries an assigned “codon pair score,” or “CPS” that is specific for a given “training set” of genes. The CPS of a given codon pair is defined as the log ratio of the observed number of occurrences over the number that would have been expected in this set of genes (in this example the human genome). Determining the actual number of occurrences of a particular codon pair (or in other words the likelihood of a particular amino acid pair being encoded by a particular codon pair) is simply a matter of counting the actual number of occurrences of a codon pair in a particular set of coding sequences. Determining the expected number, however, requires additional calculations. The expected number is calculated so as to be independent of both amino acid frequency and codon bias similarly to Gutman and Hatfield. That is, the expected frequency is calculated based on the relative proportion of the number of times an amino acid is encoded by a specific codon. A positive CPS value signifies that the given codon pair is statistically over-represented, and a negative CPS indicates the pair is statistically under-represented in the human genome.

To perform these calculations within the human context, the most recent Consensus CDS (CCDS) database of consistently annotated human coding regions, containing a total of 14,795 genes, was used. This data set provided codon and codon pair, and thus amino acid and amino-acid pair frequencies on a genomic scale.

The paradigm of Federov et al. (2002), was used to further enhanced the approach of Gutman and Hatfield (1989). This allowed calculation of the expected frequency of a given codon pair independent of codon frequency and non-random associations of neighboring codons encoding a particular amino acid pair.

$$S(P_{ij}) = \ln\left(\frac{N_O(P_{ij})}{N_E(P_{ij})}\right) = \ln\left(\frac{N_O(P_{ij})}{F(C_i)F(C_j)N_O(X_{ij})}\right)$$

In the calculation, P_{ij} is a codon pair occurring with a frequency of $N_O(P_{ij})$ in its synonymous group. C_i and C_j are the two codons comprising P_{ij} , occurring with frequencies $F(C_i)$ and $F(C_j)$ in their synonymous groups respectively. More explicitly, $F(C_i)$ is the frequency that corresponding amino acid X_i is coded by codon C_i throughout all coding regions and $F(C_i)=N_O(C_i)/N_O(X_i)$, where $N_O(C_i)$ and

$N_o(X_i)$ are the observed number of occurrences of codon C_i and amino acid X_i respectively. $F(C_i)$ is calculated accordingly. Further, $N_o(X_{ij})$ is the number of occurrences of amino acid pair X_{ij} throughout all coding regions. The codon pair bias score $S(P_{ij})$ of P_{ij} was calculated as the log-odds ratio of the observed frequency $N_o(P_{ij})$ over the expected number of occurrences of $N_e(P_{ij})$.

Using the formula above, it was then determined whether individual codon pairs in individual coding sequences are over- or under-represented when compared to the corresponding genomic $N_e(P_{ij})$ values that were calculated by using the entire human CCDS data set. This calculation resulted in positive $S(P_{ij})$ score values for over-represented and negative values for under-represented codon pairs in the human coding regions (FIG. 7).

The “combined” codon pair bias of an individual coding sequence was calculated by averaging all codon pair scores according to the following formula:

$$S(P_{ij}) = \sum_{l=1}^k \frac{S(P_{ij})}{k-1}.$$

The codon pair bias of an entire coding region is thus calculated by adding all of the individual codon pair scores comprising the region and deviding this sum by the length of the coding sequence.

Changing of Codon Pair Bias.

The capsid-coding region of PV(M) was re-engineered to change codon pair bias. The largest possible number of rarely used codon pairs (creating virus PV-Min) or the largest possible number of widely used codon pairs (creating virus PV-Max) was introduced, while preserving the codon bias and all other features of the wt virus genome. The following explains our method in detail.

Two sequences were designed to vary the poliovirus P1 region codon pair score in the positive (PV-Max; SEQ ID NO:4) and negative (PV-Min; SEQ ID NO:5) directions. By leaving the amino acid sequence unaltered and the codon bias minimally modified, a simulated annealing algorithm was used for shuffling codons, with the optimization goal of a minimum or maximum codon pair score for the P1 capsid region. The resulting sequences were processed for elimination of splice sites and reduction of localized secondary structures. These sequences were then synthesized by a commercial vendor, Blue Heron Biotechnology, and sequence-verified. The new capsid genes were used to replace the equivalent wt sequence in an infectious cDNA clone of wt PV via two PflMI restriction sites. Virus was derived as described in Example 1.

For the PV-Max virus, death of infected cells was seen after 24 h, a result similar to that obtained with wt virus. Maximal viral titer and one-step growth kinetics of PV-Max were also identical to the wt. In contrast, no cell death resulted in cells transfected with PV-Min mutant RNA and no viable virus could be recovered. The transfections were repeated multiple times with the same result. Lysates of PV-Min transfected cells were subjected to four successive blind passages, and still no virus was obtained.

The capsid region of PV-Min was divided into two smaller sub-fragments (PV-Min⁷⁵⁵⁻²⁴⁷⁰ and PV-Min²⁴⁷⁰⁻³³⁸⁵) as had been done for PV-AB (poor codon bias), and the sub-fragments were cloned into the wt background. As with the PV-AB subclones, subclones of PV-Min were very sick, but not dead (FIG. 8). As observed with PV-AB viruses, the

phenotype of PV-Min viruses is a result of reduced specific infectivity of the viral particles rather than to lower production of progeny virus. Ongoing studies involve testing the codon pair-attenuated chimeras in CD155tg mice to determine their pathogenicity. Also, additional chimeric viruses comprising subclones of PV-Min cDNAs are being made, and their ability to replicate is being determined (see example 8 and 9 below). Also, the effect of distributing intermediate amounts of codon pair bias over a longer sequence are being confirmed. For example, a poliovirus derivative is designed to have a codon pair bias of about -0.2 (PV-0.2; SEQ ID NO:6), and the mutations from wild type are distributed over the full length of the P1 capsid region. This is in contrast to PV-MinZ (PV-Min²⁴⁷⁰⁻³³⁸⁶) which has a similar codon pair bias, but with codon changes distributed over a shorter sequence.

It is worth pointing out that PV-Min and PV-0.2 are sequences in which there is little change in codon usage relative to wild type. For the most part, the sequences employ the same codons that appear in the wild type PV(M) virus. PV-MinZ is somewhat different in that it contains a portion of PV-Min subcloned into PV(M). As with PV-Min and PV-0.2, the encoded protein sequence is unchanged, but codon usage as determined in either the subcloned region, or over the entire P1 capsid region, is not identical to PV-Min (or PV-0.2), because only a portion of the codon rearranged sequence (which has identical codons over its full length, but not within smaller segments) has been substituted into the PV(M) wild type sequence. Of course, a mutated capsid sequence could be designed to have a codon pair bias over the entire P1 gene while shuffling codons only in the region from nucleotides 2470-3386.

EXAMPLE 8

Viruses Constructed by a Change of Codon-Pair Bias are Attenuated in CD155 tg Mice

Mice Intracerebral Injections, Survival

To test the attenuation of PV-Min⁷⁵⁵⁻²⁴⁷⁰ and PV-Min²⁴⁷⁰⁻³³⁸⁵ in an animal model, these viruses were purified and injected intra-cerebrally into CD 155 (PVR/poliovirus receptor) transgenic mice (See Table 5). Indeed these viruses showed a significantly attenuated phenotype due to the customization of codon pair bias using our algorithm. PVM-wt was not injected at higher dose because all mice challenged at 10e5 virions died because of PVM-wt. This attenuated phenotype is due to the customization of codon pair bias using our algorithm. This reaffirms that the customization of codon-pair bias is applicable for a means to create live vaccines.

TABLE 5

Virus	Mice Intracerebral Injections, Survival			
	10e4 Virions	10e5 Virions	10e6 Virions	10e7 Virions
PV-Min ⁷⁵⁵⁻²⁴⁷⁰	4/4	3/4	3/5	3/4
PV-Min ²⁴⁷⁰⁻³³⁸⁵	4/4	4/4	5/5	3/4
PVM-wt	3/4	0/4	—	—

These findings are significant in two respects. First, they are the first clear experimental evidence that codon pair bias is functionally important, i.e., that a deleterious phenotype can be generated by disturbing codon pair bias. Second, they provide an additional dimension of synonymous codon changes that can be used to attenuate a virus. The in vivo

pathogenicity of these codon-pair attenuated chimeras have been tested in CD155tg and have shown an attenuated phenotype (See Table 5). Additional chimeric viruses comprising subclones of PV-Min capsid cDNAs have been assayed for replication in infected cells and have also shown an attenuated phenotype.

EXAMPLE 9

Construction of Synthetic Poliovirus with Altered Codon-Pair Bias: Implications for Vaccine Development

Calculation of Codon Pair Bias, Implementation of Algorithm to Produce Codon Pair Deoptimized Sequences.

We developed an algorithm to quantify codon pair bias. Every possible individual codon pair was given a "codon pair score", or "CPS". We define the CPS as the natural log of the ratio of the observed over the expected number of occurrences of each codon pair over all human coding regions.

$$CPS = \ln \left(\frac{F(AB)^o}{\frac{F(A) \times F(B)}{F(X) \times F(Y)} \times F(XY)} \right)$$

Although the calculation of the observed occurrences of a particular codon pair is straightforward (the actual count within the gene set), the expected number of occurrences of a codon pair requires additional calculation. We calculate This expected number is calculated to be independent both of amino acid frequency and of codon bias, similar to Gutman and Hatfield. That is, the expected frequency is calculated based on the relative proportion of the number of times an amino acid is encoded by a specific codon. A positive CPS value signifies that the given codon pair is statistically over-represented, and a negative CPS indicates the pair is statistically under-represented in the human genome

Using these calculated CPSs, any coding region can then be rated as using over- or under-represented codon pairs by taking the average of the codon pair scores, thus giving a Codon Pair Bias (CPB) for the entire gene.

$$CPB = \sum_{i=1}^k \frac{CPS_i}{k-1}$$

The CPB has been calculated for all annotated human genes using the equations shown and plotted (FIG. 7). Each point in the graph corresponds to the CPB of a single human gene. The peak of the distribution has a positive codon pair bias of 0.07, which is the mean score for all annotated human genes. Also there are very few genes with a negative codon pair bias. Equations established to define and calculate CPB were then used to manipulate this bias.

Development and Implementation of Computer-Based Algorithm to Produce Codon Pair Deoptimized Sequences.

Using these formulas we next developed a computer based algorithm to manipulate the CPB of any coding region while maintaining the original amino acid sequence. The algorithm has the critical ability to maintain the codon usage of a gene (i.e. preserve the frequency of use of each existing codon) but "shuffle" the existing codons so that the CPB can be increased or decreased. The algorithm uses simulated annealing, a mathematical process suitable for full-length

optimization (Park, S. et al., 2004). Other parameters are also under the control of this algorithm; for instance, the free energy of the folding of the RNA. This free energy is maintained within a narrow range, to prevent large changes in secondary structure as a consequence of codon re-arrangement. The optimization process specifically excludes the creation of any regions with large secondary structures, such as hairpins or stem loops, which could otherwise arise in the customized RNA. Using this computer software the user simply needs to input the cDNA sequence of a given gene and the CPB of the gene can be customized as the experimenter sees fit.

De Novo Synthesis of P1 Encoded by Either Over-Represented or Under-Represented Codon-Pairs.

To obtain novel, synthetic poliovirus with its P1 encoded by either over-represented or under-represented codon pairs, we entered the DNA sequence corresponding to the P1 structural region of poliovirus type I Mahoney (PV(M)-wt) into our program yielding-PV-Max-P1 using over-represented codon pairs (566 mutations) and PV-Min-P1 using under-represented codon pairs (631 mutations). The CPB scores of these customized, novel synthetic P-1 regions are PV-Max=+0.25 and PV-Min=-0.48, whereas the CPB of PV(M)-wt is -0.02 (FIG. 7).

Additional customization included inclusion of restriction sites that were designed into both synthetic sequences at given intervals, to allow for sub-cloning of the P1 region. These synthetic P1 fragments were synthesized de novo by Blue Heron Corp. and incorporated into a full-length cDNA construct of poliovirus (FIG. 11) (Karlin et al., 1994). A small fragment (3 codons, 9 nucleotides) of PV(M)-wt sequence was left after the AUG start codon in both constructs to allow translation to initiate equally for all synthetic viruses; thus providing more accurate measurement of the effect of CPB on the elongation phase of translation.

DNA Synthesis, Plasmids, Sub Cloning of Synthetic Capsids and Bacteria.

Large codon-pair altered PV cDNA fragments, corresponding to nucleotides 495 to 3636 of the PV genome, were synthesized by Blue Heron Corp. using their proprietary GeneMaker® system (<http://www.blueheronbio.com/>). All subsequent poliovirus cDNA clones/sub clones were constructed from PV1(M) cDNA clone pT7PVM using unique restriction sites (van der Wert, et al., 1986). The full-length PV-Min, PV-Max cassette was released from Blue Heron's carrier vector via PflMI digestion and insertion into the pT7PVM vector with its PflMI fragment removed. The PV-MinXY and PV-MinZ constructs were obtained by digestion with NheI and BglII simultaneously, then swapping this fragment with a pT7PVM vector digested similarly. PV-MinXY and PV-MinZ were constructed via BsmI digestion and exchanging the fragment/vector with the similarly digested pT7PVM. PV-MinY was constructed by digesting the PV-MinXY construct with BsmI and swapping this fragment with the BsmI fragment for a digested pT7PVM. Plasmid transformation and amplification were all achieved via *Escherichia coli* DH5α.

Creation of Chimeric Viruses Containing CPB-Altered Capsid Regions: Under-Represented Codon Pair Bias throughout the P1 Results in a Null Phenotype.

Using the T7 RNA polymerase promoter upstream of the poliovirus genomic sequence, positive-sense RNA was transcribed. 1.5 µg of a given plasmid cDNA clone from above was linearized via an EcoRI digestion and then was transcribed into RNA via T7 RNA polymerase (Stratagene) driven by its promoter upstream of the cDNA for 2 hours at 37°C. (van der Werf et al., 1986). This RNA was transfected

into 1×10^6 HeLa R19 cells using a modified DEAE-Dextran method (van der Werf et al., 1986). These cells were than incubate at room-temperature (RT) for 30-minutes. The transfection supernatant was removed and Dulbecco's modified Eagle medium (DMEM) containing 2% bovine calf serum (BCS) was added and the cells were incubated at 37° C. and observed (up to 4 days) for the onset of cytopathic effect (CPE).

The PV-Max RNA transfection produced 90% cytopathic effect (CPE) in 24 hours, which is comparable to the transfection of PV(M)-wt RNA. The PV-Max virus generated plaques identical in size to the wild type. In contrast, the PV-Min RNA produced no visible cytopathic effect after 96 hours, and no viable virus could be isolated even after four blind passages of the supernatant from transfected cells.

The subsequent use of the supernatant from cells subjected to PV-Max RNA transfection also produced 95% CPE in 12 hours, thus indicating that the transfected genomic material successfully produced PV-Max poliovirus virions. In contrast, the PV-Min viral RNA yielded no visible CPE after 96 hours and four blind passages of the supernatant, possibly containing extremely low levels of virus, also did not produce CPE. Therefore the full-length PV-Min synthetic sequence, utilizing under-represented codon pairs, in the P1 region cannot generate viable virus and so it would need to be sub-cloned.

HeLa R19 cells were maintained as a monolayer in DMEM containing 10% BCS. Virus amplification was achieved on (1.0×10^8 cells) HeLa R19 monolayers using 1 M.O.I. Infected cells were incubated at 37° C. in DMEM with 2% BCS for three days or until CPE was observed. After three freeze/thaw cycles cell debris was removed form the lysates via low speed centrifugation and the supernatant containing virus was used for further experiments.

One-Step growth curves were achieved by infecting a monolayer of HeLa R19 cells with 5 M.O.I. of a given virus, the inoculum was removed, cells washed 2x with PBS and then incubating at 37° C. for 0, 2, 4, 7, 10, 24, and 48 hours. These time points were then analyzed via plaque assay. All Plaque assay were performed on monolayers of HeLa R19 cells. These cells were infected with serial dilution of a given growth curve time point or purified virus. These cells were then overlaid with a 0.6% tragenthum gum in Modified Eagle Medium containing 2% BCS and then incubated at 37° C. for either 2 days for PV(M)-wt and PV-Max, or 3 days for PV-Min (X, Y, XY, or Z) viruses. These were then developed via crystal violet staining and the PFU/ml titer was calculated by counting visible plaques.

Small Regions of Under-Represented Codon Pair Bias Rescues Viability, but Attenuate the Virus.

Using the restriction sites designed within the PV-Min sequence we subcloned portions of the PV-Min P1 region into an otherwise wild-type virus, producing chimeric viruses where only sub-regions of P1 had poor codon pair bias (FIG. 11) (van der Werf et al., 1986). From each of these sub-clones, RNA was produced via *in vitro* transcription and then transfected into HeLa R19 cells, yielding viruses with varying degrees of attenuation (Viability scores, FIG. 11). P1 fragments X and Y are each slightly attenuated; however when added together they yield a virus (PV-Min⁷⁵⁵⁻²⁴⁷⁰, PV-MinXY) that is substantially attenuated (FIGS. 3, 4). Virus PVMin²⁴⁷⁰⁻³³⁸⁵ (PV-MinZ) is about as attenuated as PV-MinXY. Construct PV-Min¹⁵¹³⁻³³⁸⁵ (YZ) did not yield plaques, and so apparently is too attenuated to yield viable virus. These virus constructs, which displayed varying degrees of attenuation were further investigated to determine their actual growth kinetics.

One-Step Growth Kinetics and the Mechanism of Attenuation: Specific Infectivity is Reduced.

For each viable construct, one step-growth kinetics were examined. These kinetics are generally similar to that of wild-type in that they proceed in the same basic manner (i.e. an eclipse phase followed by rapid, logarithmic growth). However, for all PV-Min constructs, the final titer in terms of Plaque Forming Units (PFU) was typically lower than that of wild-type viruses by one to three orders of magnitude (FIG. 12A).

When virus is measured in viral particles per ml (FIG. 12B) instead of PFU, a slightly different result is obtained and suggests these viruses produce nearly equivalent numbers of particles per cell per cycle of infection as the wild-type virus. In terms of viral particles per ml, the most attenuated viruses are only 78% (PV-MinXY) or 82% (PV-MinZ) attenuated which on a log scale is less than one order of magnitude. Thus these viruses appear to be attenuated by about two orders of magnitude in their specific infectivity (the number of virions required to generate a plaque).

To confirm that specific infectivity was reduced, we re-measured the ratio of viral particles per PFU using highly purified virus particles. Selected viruses were amplified on 10^8 HeLa R19 cells. Viral lysates were treated with RNase A to destroy exposed viral genomes and any cellular RNAs, that would obscure OD values. Also the viral lysates were then incubated for 1 hour with 0.2% SDS and 2mM EDTA to denature cellular and non-virion viral proteins. A properly folded and formed poliovirus capsid survives this harsh SDS treatment, were as alph particles do not (Mueller et al., 2005). Virions from these treated lysates were then purified via ultracentrifugation over a sucrose gradient. The virus particle concentration was measured by optical density at 260 nm using the formula 9.4×10^{12} particles/ml = 1 OD₂₆₀ unit (Rueckert, 1985). A similar number of particles was produced for each of the four viruses (Table 6). A plaque assay was then performed using these purified virions. Again, PV-MinXY and PV-MinZ required many more viral particles than wild-type to generate a plaque (Table 6).

For wild-type virus, the specific infectivity was calculated to be 1 PFU per 137 particles (Table 6), consistent with the literature (Mueller et al., 2006; Schwerdt and Fogh, 1957; Joklik and Darnell, 1961). The specific infectivities of viruses PV-MinXY and PV-MinZ are in the vicinity of 1 PFU per 10,000 particles (Table 6).

Additionally the heat stability of the synthetic viruses was compared to that of PV(M)-wt to reaffirm the SDS treatment data, that these particles with portions of novel RNA were equally as stable. Indeed these synthetic viruses had the same temperature profile as PV(M)-wt when incubated at 50° C. and quantified as a time course (data not shown).

Under-Represented Codon Pairs Reduce Translation Efficiency, whereas Over-Represented Pairs Enhance Translation.

One hypothesis for the existence of codon pair bias is that the utilization of under-represented pairs causes poor or slow translation rates. Our synthetic viruses are, to our knowledge, the first molecules containing a high concentration of under-represented codon pairs, and as such are the first molecules suitable for a test of the translation hypothesis.

To measure the effect of codon pair bias on translation, we used a dicistronic reporter (Mueller et al., 2006) (FIG. 13). The first cistron expresses *Renilla luciferase* (R-Luc) under the control of the hepatitis C virus internal ribosome entry site (IRES) and is used as a normalization control. The second cistron expresses firefly luciferase (F-Luc) under the

control of the poliovirus IRES. However, in this second cistron, the F-Luc is preceded by the P1 region of poliovirus, and this P1 region could be encoded by any of the synthetic sequence variants described here. Because F-Luc is translated as a fusion protein with the proteins of the P1 region, the translatability of the P1 region directly affects the amount of F-Luc protein produced. Thus the ratio of F-Luc luminescence to R-Luc luminescence is a measure of the translatability of the various P1 encodings.

The P1 regions of wild-type, PV-Max, PV-Min, PV-MinXY and PV-MinZ were inserted into the region labeled "P1" (FIG. 13A). PV-MinXY, PV-MinZ, and PV-Min produce much less F-Luc per unit of R-Luc than does the wild-type P1 region, strongly suggesting that the under-represented codon pairs are causing poor or slow translation rates (FIG. 13). In contrast, PV-Max P1 (which uses over-represented codon pairs) produced more F-Luc per unit of R-Luc, suggesting translation is actually better for PV-Max P1 compared to PV(M)-wt P1.

Dicistronic Reporter Construction, and In Vivo Translation.

The dicistronic reporter constructs were all constructed based upon pdiLuc-PV (Mueller et al., 2006). PV-Max and PV-Min capsid regions were amplified via PCR using the oligonucleotides P1max-2A-RI (+)/P1max-2A-RI (-) or P1min-2A-RI (+)/P1min-2A-RI (-) respectively. The PCR fragment was gel purified and then inserted into an intermediate vector pCR- \emptyset -XL-TOPO \emptyset (Invitrogen). This intermediate vector was then amplified in One Shot \emptyset TOP10 chemically competent cells. After preparation of the plasmid via Quiagene miniprep the intermediate vectors containing PV-Min was digested with EcoRI and these fragments were ligated into the pdiLuc-PV vector that was equally digested with EcoRI (Mueller et al., 2006). These plasmids were also amplified in One Shot \emptyset TOP10 chemically competent cells (Invitrogen). To construct pdiLuc-PV-MinXY and pdiLuc-PV-MinZ, pdiLuc-PV and pdiLuc-PV-Min were equally digested with NheI and the resulting restriction fragments were exchanged between the respective vectors. These were then transformed into One Shot \emptyset TOP10 chemically competent cells and then amplified. From all four of these clones RNA was transcribed via the T7 polymerase method (van der Werf et al., 1986).

To analyze the in vivo translation efficiency of the synthetic capsids the RNA of the dicistronic reporter constructs were transfected into 2×10^5 HeLa R19 cells on 12-well dishes via Lipofectamine 2000 (Invitrogen). In order to quantify the translation of only input RNA the transfection was accomplished in the presence of 2 mM guanidine hydrochloride (GuHCl). Six hours after transfection cells were lysed via passive lysis buffer (Promega) and then these lysates were analyzed by a dual firefly (F-Luc) *Renilla* (R-Luc) luciferase assay (Promega).

Genetic Stability of PV-MinXY and PV-MinZ.

Because PV-MinXY and PV-MinZ each contain hundreds of mutations (407 and 224, respectively), with each mutation causing a minuscule decrease in overall codon pair bias, we believe it should be very difficult for these viruses to revert to wild-type virulence. As a direct test of this idea, viruses PV-MinXY and PV-MinZ were serially-passaged 15 times, respectively, at an MOI of 0.5. The titer was monitored for phenotypic reversion, and the sequence of the passaged virus was monitored for reversions or mutation. After 15 passages there was no phenotypic change in the viruses (i.e.

same titer, induction of CPE) and there were no fixed mutations in the synthetic region.

Heat Stability and Passaging.

The stability of the synthetic viruses, PV-MinXY and PV-Min Z, was tested and compared to PV(M)-wt. This was achieved by heating 1×10^8 particles suspended in PBS to 50° C. for 60 minutes and then measuring the decrease in intact viral particles via plaque assay at 5, 15, 30 and 60 minutes (FIG. 14). In order to test the genetic stability of the synthetic portions of the P1 region of the viruses PV-MinXY and PV-MinZ these viruses were serial passaged. This was achieved by infecting a monolayer of 1×10^6 HeLa R19 cells with 0.5 MOI of viruses, PV-MinXY and PV-MinZ, and then waiting for the induction of CPE. Once CPE initiated, which remained constant throughout passages, the lysates were used to infect new monolayers of HeLa R19 cells. The titer and sequence was monitored at passages 5, 9, and 15 (data not shown).

Virus Purification and Determination of Viral Particles via OD₂₆₀ Absorbance.

A monolayer of HeLa R19 cells on a 15 cm dish (1×10^8 cells) were infected with PV(M)-wt, PV-Max, PV-MinXY or PV-Min Z until CPE was observed. After three freeze/thaw cycles the cell lysates were subjected to two initial centrifugations at 3,000 \times g for 15 minutes and then 10,000 \times g for 15 minutes. Then 10 μ g/ml of RNase A (Roche) was added to supernatant and incubated at RT for 1 hour; Subsequently 0.5% sodium dodecyl sulfate (SDS) and 2 mM EDTA was added to the supernatant, gently mixed and incubated at RT for 30 minutes. These supernatants containing virus particles were placed above a 6 ml sucrose cushion [30% sucrose in Hank's Buffered Salt Solution (HBSS)]. Sedimentation of virus particles was achieved by ultracentrifugation through the sucrose gradient for 3.5 hours at 28,000 rpm using an SW28 swing-bucket rotor.

After centrifugation, the sucrose cushion was left intact and the supernatant was removed and the tube was washed two times with HBBS. After washing, the sucrose was removed and the virus "pearl" was re-suspended in PBS containing 0.1% SDS. Viral titers were determined via plaque assay (above). Virus particles concentration was determined via the average of three measurements of the optical density at 260 nm of the solution via the NanoDrop spectrophotometer (NanoDrop Technologies) using the formula 9.4×10^{12} particles/ml = 1 OD₂₆₀ unit (Mueller et al., 2006; Rueckert, 1985).

Neuroattenuation of PV-MinXY and PV-MinZ in CD155tg Mice.

The primary site of infection of wild-type poliovirus is the oropharynx and gut, but this infection is relatively asymptomatic. However, when the infection spreads to motor neurons in the CNS in 1% of PV(M)-wt infections, the virus destroys these neurons, causing death or acute flaccid paralysis known as poliomyelitis (Landsteiner and Popper, 1909; Mueller et al., 2005). Since motor neurons and the CNS are the critical targets of poliovirus, we wished to know whether the synthetic viruses were attenuated in these tissues. Therefore these viruses were administered to CD155tg mice (transgenic mice expressing the poliovirus receptor) via intracerebral injection (Koike et al., 1991). The PLD₅₀ value was calculated for the respective viruses and the PV-MinXY and PV-MinZ viruses were attenuated either 1,000 fold based on particles or 10 fold based on PFU (Table 6) (Reed and Muench, 1938). Since these viruses did display neuroattenuation they could be used as a possible vaccine.

TABLE 6

Virus	A ₂₆₀	Purified Particles/ml ^a	Purified PFU/ml	Specific Infectivity ^b	PLD ₅₀ (Particles) ^c	PLD ₅₀ (PFU) ^d
PV-M(wt)	0.956	8.97 × 10 ¹²	6.0 × 10 ¹⁰	1/137	10 ^{4.0}	10 ^{1.9}
PV-Max	0.842	7.92 × 10 ¹²	6.0 × 10 ¹⁰	1/132	10 ^{4.1}	10 ^{1.9}
PV-MinXY	0.944	8.87 × 10 ¹²	9.6 × 10 ⁸	1/9,200	10 ^{7.1}	10 ^{3.2}
PV-MinZ	0.731	6.87 × 10 ¹²	5.1 × 10 ⁸	1/13,500	10 ^{7.3}	10 ^{3.2}

^aThe A₂₆₀ was used to determine particles/ml via the formula 9.4 × 10¹² particles/ml = 1 OD₂₆₀ unit^bCalculated by dividing the PFU/ml of purified virus by the Particles/ml^{c,d}calculated after administration of virus via intracerebral injection to CD155tg mice at varying doses

Vaccination of CD155tg Mice Provides Immunity and
Protection against Lethal Challenge.

Groupings of 4-6, 6-8 week old CD155tg mice (Tg21 strain) were injected intracerebrally with purified virus dilutions from 10² particles to 10⁹ particles in 30 ul PBS to determine neuropathogenicity (Koike, et al., 1991).

The lethal dose (LD₅₀) was calculated by the Reed and Muench method (Reed and Muench, 1938). Viral titers in the spinal chord and brain were quantified by plaque assay (data not shown).

PV-MinZ and PV-MinXY encode exactly the same proteins as wild-type virus, but are attenuated in several respects, both a reduced specific infectivity and neuroattenuation.

To test PV-Min Z, PV-MinXY as a vaccine, three sub-lethal dose (10⁸ particles) of this virus was administered in 100 ul of PBS to 8, 6-8 week old CD155tg mice via intraperitoneal injection once a week for three weeks. One mouse from the vaccine cohort did not complete vaccine regimen due to illness. Also a set of control mice received three mock vaccinations with 100 ul PBS. Approximately one week after the final vaccination, 30 ul of blood was extracted from the tail vein. This blood was subjected to low speed centrifugation and serum harvested. Serum conversion against PV(M)-wt was analyzed via micro-neutralization assay with 100 plaque forming units (PFU) of challenge virus, performed according to the recommendations of WHO (Toyoda et al., 2007; Wahby, A. F., 2000). Two weeks after the final vaccination the vaccinated and control mice were challenged with a lethal dose of PV(M)-wt by intramuscular injection with a 10⁶ PFU in 100 ul of PBS (Toyoda et al., 2007). All experiments utilizing CD155tg mice were undertaken in compliance with Stony Brook University's IACUC regulations as well as federal guidelines. All 14 vaccinated mice survived and showed no signs of paralysis or parasympathetic; in contrast, all mock-vaccinated mice died (Table 7). These data suggest that indeed the CPB virus using de-optimized codon pairs is able to immunize against the wild-type virus, providing both a robust humoral response, and also allowing complete survival following challenge.

TABLE 7

Protection Against Lethal Challenge	
Virus ^a	Mice Protected (out of 7) ^b
PV-MinZ	7
PV-MinXY	7
Mock vaccinated	0

^aCD155tg mice received three vaccination doses (10⁸ particles) of respective virus^bchallenged with 10⁶ PFU of PV(M)-wt via intramuscular injection.

EXAMPLE 10

Application of SAVe to Influenza Virus

Influenza virus has 8 separate genomic segments. GenBank deposits disclosing the segment sequences for Influenza A virus (A/Puerto Rico/8/34/Mount Sinai(H1N1)) include AF389115 (segment 1, Polymerase PB2), AF389116 (segment 2, Polymerase PB1), AF389117 (segment 3, Polymerase PA), AF389118 (segment 4, hemagglutinin HA), AF389119 (segment 5, nucleoprotein NP), AF389120 (segment 6, neuraminidase NA), AF389121 (segment 7, matrix proteins M1 and M2), and AF389122 (segment 8, nonstructural protein NS1).

In initial studies, the genomic segment of strain A/PR/8/34 (also referred to herein as A/PR8) encoding the nucleoprotein NP, a major structural protein and the second most abundant protein of the virion (1,000 copies per particle) that binds as monomer to full-length viral RNAs to form coiled ribonucleoprotein, was chosen for deoptimization. (See Table 8, below, for parent and deoptimized sequences). Moreover, NP is involved in the crucial switch from mRNA to template and virion RNA synthesis (Palese and Shaw, 2007). Two synonymous encodings were synthesized, the first replacing frequently used codons with rare synonymous codons (NP^{CD}) (i.e., de-optimized codon bias) and, the second, de-optimizing codon pairs (NP^{CPmin}). The terminal 120 nucleotides at either end of the segment were not altered so as not to interfere with replication and encapsidation. NP^{CD} contains 338 silent mutations and NP^{CPmin} (SEQ ID NO:23) contains 314 silent mutations. The mutant NP segments were introduced into ambisense vectors as described (below), and together with the other seven wt influenza plasmids co-transfected into 293T/MDCK co-cultured cells. As a control, cells were transfected with all 8 wt A/PR8 plasmids. Cells transfected with the NP^{CD} segment and the NP^{CPmin} segment produced viable influenza virus similarly to cells transfected with wild-type NP. These new de-optimized viruses, referred to as A/PR8-NP^{CD} or A/PR8-NP^{CPmin}, respectively, appear to be attenuated: The titer (in terms of PFU) is 3- to 10-fold lower than the wild-type virus, and the mutant viruses both make small plaques.

Although the de-optimized influenza viruses are not as severely attenuated as a poliovirus containing a similar number of de-optimized codons, there is a difference in the translational strategies of the two viruses. Poliovirus has a single long mRNA, translated into a single polyprotein. Slow translation through the beginning of this long mRNA (as in our capsid de-optimized viruses) will reduce translation of the entire message, and thus affect all proteins. In contrast, influenza has eight separate segments, and deoptimization of one will have little if any effect on transla-

tion of the others. Moreover, expression of the NP protein is particularly favored early in influenza virus infection (Palese and Shaw, 2007).

Characterization of Influenza Virus Carrying a Codon Pair Deoptimized NP Segment

The growth characteristics of A/PR8-NP^{CPmin} were analyzed by infecting confluent monolayers of Madin Darby Canine Kidney cells (MDCK cells) in 100 mm dishes with 0.001 multiplicities of infection (MOI). Virus inoculums were allowed to adsorb at room temperature for 30 minutes on a rocking platform, then supplemented with 10 ml of Dulbecco Modified Eagle Medium (DMEM) containing 0.2% Bovine Serum Albumin (BSA) and 2 µg/ml TPCK treated Trypsin and incubated at 37 °C. After 0, 3, 6, 9, 12, 24, and 48 hours, 100 µl of virus containing medium was removed and virus titers determined by plaque assay.

Viral titers and plaque phenotypes were determined by plaque assay on confluent monolayers of MDCK cells in 35 mm six well plates. 10-fold serial dilutions of virus were prepared in Dulbecco Modified Eagle Medium (DMEM) containing 0.2% Bovine Serum Albumin (BSA) and 2 µg/ml TPCK treated Trypsin. Virus dilutions were plated out on MDCK cells and allowed to adsorb at room temperature for 30 minutes on a rocking platform, followed by a one hour incubation at 37 °C in a cell culture incubator. The inoculum was then removed and 3 ml of Minimal Eagle Medium containing 0.6% tragacanth gum (Sigma-Aldrich) 0.2% BSA and 2 µg/ml TPCK treated Trypsin. After 72 hours of incubation at 37 °C, plaques were visualized by staining the wells with crystal violet.

A/PR8-NP^{Min} produced viable virus that produced smaller plaques on MDCK cells compared to the A/PR8 wt (FIG. 16A). Furthermore, upon low MOI infection A/PR8-NP^{Min} manifests a delayed growth kinetics, between 3-12 hrs post infection, where A/PR8-NP^{Min} titers lags 1.5 logs behind A/PR8 (FIG. 16B). Final titers are were 3-5 fold lower than that of A/PR8 (average of three different experiments).

Characterization of Influenza Viruses A/PR8-PB1^{Min-RR}, A/PR8-HA^{Min} and A/PR8-HA^{Min}/NP^{Min} Carrying Codon Pair Deoptimized PB1, HA, or HA and NP Segments.

Codon pair de-optimized genomic segments of strain A/PR/8/34 encoding the hemagglutinin protein HA and the polymerase subunit PB1 were produced. HA is a viral structural protein protruding from the viral surface mediating receptor attachment and virus entry. PB1 is a crucial component of the viral RNA replication machinery. Specifically a synonymous encoding of PB1 (SEQ ID NO:15) was synthesized by de-optimizing codon pairs between codons 190-488 (nucleotides 531-1488 of the PB1 segment) while retaining the wildtype codon usage (PB1^{Min-RR}). Segment PB1^{Min-RR} contains 236 silent mutations compared the wt PB1 segment.

A second synonymous encoding of HA (SEQ ID NO:21) was synthesized by de-optimizing codon pairs between codons 50-541 (nucleotides 180-1655 of the HA segment) while retaining the wildtype codon usage (HA^{Min}). HA^{Min} contains 355 silent mutations compared the to wt PB1 segment.

The mutant PB1^{Min-RR} and HA^{Min} segments were introduced into an ambisense vector as described above and together with the other seven wt influenza plasmids co-transfected into 293T/MDCK co-cultured cells. In addition the HA^{Min} segment together with the NP^{Min} segment and the

remaining six wt plasmids were co-transfected. As a control, cells were transfected with all 8 wt A/PR8 plasmids. Cells transfected with either PB1^{Min-RR} or HA^{Min} segments produced viable virus as did the combination of the codon pair deoptimized segments HA^{Min} and NP^{Min}. The new deoptimized viruses are referred to as A/PR8-PB1^{Min-RR}, A/PR8-HA^{Min}, and A/PR8-HA^{Min}/NP^{Min}, respectively.

Growth characteristics and plaque phenotypes were assessed as described above.

A/PR8-PB1^{Min-RR}, A/PR8-HA^{Min}, and A/PR8-HA^{Min}/NP^{Min} all produced viable virus. A/PR8-PB1^{Min-RR} and A/PR8-HA^{Min}/NP^{Min} produced smaller plaques on MDCK cells compared to the A/PR8 wt (FIG. 17A). Furthermore, upon low MOI infection on MDCK cells A/PR8-HA^{Min} and A/PR8-HA^{Min}/NP^{Min} display much reduced growth kinetics, especially from 3-12 hrs post infection, where A/PR8-HA^{Min}/NP^{Min} titers lag 1 to 2 orders of magnitude behind A/PR8 (FIG. 17B). Final titers for both A/PR8-HA^{Min} and A/PR8-HA^{Min}/NP^{Min} were 10 fold lower than that of A/PR8. As A/PR8-HA^{Min}/NP^{Min} is more severely growth retarded than A/PR8-HA^{Min}, it can be concluded that the effect of deoptimizing two segments is additive.

Attenuation of A/PR8-NP^{Min} in a BALB/c Mouse Model

Groups of 6-8 anesthetized BALB/c mice 6 weeks of age were given 12.5 µl of A/PR8 or A/PR8-NP^{Min} virus solution to each nostril containing 10-fold serial dilutions between 10² and 10⁶ PFU of virus. Mortality and morbidity (weight loss, reduced activity, death) was monitored. The lethal dose 50, LD₅₀, was calculated by the method of Reed and Muench (Reed, L. J., and M. Muench. 1938. Am. J. Hyg. 27:493-497).

Eight mice were vaccinated once by intranasal inoculation with 10² PFU of A/PR8-NP^{Min} virus. A control group of 6 mice was not vaccinated with any virus (mock). 28 days following this initial vaccination the mice were challenged with a lethal dose of the wt virus A/PR8 corresponding to 100 times the LD₅₀.

The LD₅₀ for A/PR8 was 4.6×10¹ PFU while the LD₅₀ for A/PR8-NP^{Min} was 1×10³ PFU. At a dose of 10² all A/PR8-NP^{Min} infected mice survived. It can be concluded that A/PR8-NP^{Min} is attenuated in mice by more than 10 fold compared to the wt A/PR8 virus. This concentration was thus chosen for vaccination experiments. Vaccination of mice with 10² A/PR8-NP^{Min} resulted in a mild and brief illness, as indicated by a relative weight loss of less than 10% (FIG. 18A). All 8 out of 8 vaccinated mice survived. Mice infected with A/PR8 at the same dose experienced rapid weight loss with severe disease. 6 of 8 mice infected with A/PR8 died between 10 and 13 days post infection (FIG. 18B). Two mice survived and recovered from the wildtype infection.

Upon challenge with 100 times LD₅₀ of wt virus, all A/PR8-NP^{Min} vaccinated were protected, and survived the challenge without disease symptoms or weight loss (FIG. 18C). Mock vaccinated mice on the other hand showed severe symptoms, and succumbed to the infection between 9 and 11 days after challenge. It can be concluded that A/PR8-NP^{Min} induced protective immunity in mice and, thus, has potential as a live attenuated influenza vaccine. Other viruses such as A/PR8-PB1^{Min-RR} and A/PR8-HA^{Min}/NP^{Min}, yet to be tested in mice, may lead to improve further the beneficial properties of codon-pair deoptimized influenza viruses as vaccines.

77

EXAMPLE 11

Development of Higher-Throughput Methods for Making and Characterizing Viral Chimeras

Constructing Chimeric Viruses by Overlapping PCR

The “scan” through each attenuated mutant virus is performed by placing approximately 300-bp fragments from each mutant virus into a wt context using overlap PCR. Any given 300-bp segment overlaps the preceding segment by ~200 bp, i.e., the scanning window is ~300 bp long, but moves forward by ~100 bp for each new chimeric virus. Thus, to scan through one mutant virus (where only the ~3000 bp of the capsid region has been altered) requires about 30 chimeric viruses. The scan is performed in 96-well dish format which has more than sufficient capacity to analyze two viruses simultaneously.

The starting material is picogram amounts of two plasmids, one containing the sequence of the wt virus, and the other the sequence of the mutant virus. The plasmids include all the necessary elements for the PV reverse genetics system (van der Werf et al., 1986), including the T7 RNA polymerase promoter, the hammerhead ribozyme (Herold and Aldino, 2000), and the DNA-encoded poly(A) tail. Three pairs of PCR primers are used, the A, M (for Mutant), and B pairs. See FIG. 9. The M pair amplifies the desired 300 bp segment of the mutant virus; it does not amplify wt, because the M primer pairs are designed based on sequences that have been significantly altered in the mutant. The A and B pairs amplify the desired flanks of the wt viral genome. Importantly, about 20-25 bp of overlap sequence is built into the 5' ends of each M primer as well as A2 and B1, respectively; these 20-25 bps overlap (100% complementarity) with the 3' end of the A segment and the 5' end of the B segment, respectively.

To carry out the overlapping PCR, one 96-well dish contains wt plasmid DNA, and the 30 different A and B pairs in 30 different wells. A separate but matching 96-well plate contains mutant plasmid DNA and the 30 different M primer pairs. PCR is carried out with a highly processive, low error rate, heat-stable polymerase. After the first round of PCR, each reaction is treated with DpnI, which destroys the template plasmid by cutting at methylated GmATC sites. An aliquot from each wt and matching mutant reaction is then mixed in PCR reaction buffer in a third 96-well dish. This time, primers flanking the entire construct are used (i.e., the A1 and B2 primers). Since each segment (A, M, and B) is designed to overlap each adjacent segment by at least 20 bp, and since the reaction is being driven by primers that can only amplify a full-length product, the segments anneal and mutually extend, yielding full-length product after two or three cycles. This is a “3-tube” (three 96-well dish) design that may be compacted to a “1-tube” (one 96-well dish) design.

Characterization of Chimeric Viruses

Upon incubation with T7 RNA polymerase, the full length linear chimeric DNA genomes produced above with all needed upstream and downstream regulatory elements yields active viral RNA, which produces viral particles upon incubation in HeLa S10 cell extract (Molla et al., 1991) or upon transfection into HeLa cells. Alternatively, it is possible to transfet the DNA constructs directly into HeLa cells expressing the T7 RNA polymerase in the cytoplasm.

The functionality of each chimeric virus is then assayed using a variety of relatively high-throughput assays, including visual inspection of the cells to assess virus-induced CPE in 96-well format; estimation of virus production using an ELISA; quantitative measurement of growth kinetics of

78

equal amounts of viral particles inoculated into cells in a series of 96-well plates; and measurement of specific infectivity(infectious units/particle [IU/P] ratio).

The functionality of each chimeric virus can then be assayed. Numerous relatively high-throughput assays are available. A first assay may be to visually inspect the cells using a microscope to look for virus-induced CPE (cell death) in 96-well format. This can also be run an automated 96-well assay using a vital dye, but visual inspection of a 96-well plate for CPE requires less than an hour of hands-on time, which is fast enough for most purposes.

Second, 3 to 4 days after transfection, virus production may be assayed using the ELISA method described in Example 3. Alternatively, the particle titer is determined using sandwich ELISA with capsid-specific antibodies. These assays allow the identification of non-viable constructs (no viral particles), poorly replicating constructs (few particles), and efficiently replicating constructs (many particles), and quantification of these effects.

Third, for a more quantitative determination, equal amounts of viral particles as determined above are inoculated into a series of fresh 96-well plates for measuring growth kinetics. At various times (0, 2, 4, 6, 8, 12, 24, 48, 72 h after infection), one 96-well plate is removed and subjected to cycles of freeze-thawing to liberate cell-associated virus. The number of viral particles produced from each construct at each time is determined by ELISA as above.

Fourth, the IU/P ratio can be measured (see Example 3).

Higher Resolution Scans

If the lethality of the viruses is due to many small defects spread through the capsid region, as the preliminary data indicate, then many or most of the chimeras are sick and only a few are non-viable. If this is the case, higher-resolution scans are probably not necessary. Conversely, if one or more of the 300 bp segments do cause lethality (as is possible for the codon-deoptimized virus in the segment between 1513 and 2470 which, as described below, may carry a translation frameshift signal that contribute to the strong phenotype of this segment), the genome scan is repeated at higher resolution, for instance a 30 bp window moving 10 bp between constructs over the 300-bp segment, followed by phenotypic analysis. A 30-bp scan does not involve PCR of the mutant virus; instead, the altered 30-bp segment is designed directly into PCR primers for the wt virus. This allows the changes responsible for lethality to be pinpointed.

EXAMPLE 12

Ongoing Investigations into the Molecular Mechanisms Underlying SAVE
Choice of Chimeras

Two to four example chimeras from each of the two parental inviable viruses (i.e., 4 to 8 total viruses) are used in the following experiments. Viable chimeras having relatively small segments of mutant DNA, but having strong phenotypes are selected. For instance, viruses PV-AB⁷⁵⁵⁻¹⁵¹³, PV-AB²⁴⁷⁰⁻²⁹⁵⁴ and PV-AB²⁹⁵⁴⁻³³⁸⁶ from the deoptimized codon virus (see Example 1), and PV-Min⁷⁵⁵⁻²⁴⁷⁰ and PV-Min²⁴⁷⁰⁻³³⁸⁶ (see Example 7), are suitable. Even better starting chimeras, with smaller inserts that will make analysis easier, may also be obtained from the experiments described above (Example 8).

RNA Abundance/Stability

Conceivably the altered genome sequence destabilizes the viral RNA. Such destabilization could be a direct effect of

the novel sequence, or an indirect effect of a pause in translation, or other defect in translation (see, e.g., Doma and Parker, 2006). The abundance of the mutant viral RNA is therefore examined. Equal amounts of RNA from chimeric mutant virus, and wt virus are mixed and transfected into HeLa cells. Samples are taken after 2, 4, 8, and 12 h, and analyzed by Northern blotting or quantitative PCR for the two different viral RNAs, which are easily distinguishable since there are hundreds of nucleotide differences. A control with wt viral RNA compared to PV-SD (the codon-shuffled virus with a wt phenotype) is also done. A reduced ratio of mutant to wt virus RNA indicates that the chimera has a destabilized RNA.

In Vitro Translation

Translation was shown to be reduced for the codon-deoptimized virus and some of its derivatives. See Example 5. In vitro translation experiments are repeated with the codon pair-deoptimized virus (PV-Min) and its chosen chimeras. There is currently no good theory, much less any evidence, as to why deoptimized codon pairs should lead to viral inviability, and hence, investigating the effect on translation may help illuminate the underlying mechanism.

In vitro translations were performed in two kinds of extracts in Example 5. One was a “souped up” extract (Molla et al., 1991), in which even the codon-deoptimized viruses gave apparently good translation. The other was an extract more closely approximating normal *in vivo* conditions, in which the deoptimized-codon viruses were inefficiently translated. There were four differences between the extracts: the more “native” extract was not dialyzed; endogenous cellular mRNAs were not destroyed with micrococcal nuclease; the extract was not supplemented with exogenous amino acids; and the extract was not supplemented with exogenous tRNA. In the present study, these four parameters are altered one at a time (or in pairs, as necessary) to see which have the most significant effect on translation. For instance, a finding that it is the addition of amino acids and tRNA that allows translation of the codon-deoptimized virus strongly supports the hypothesis that translation is inefficient simply because rare aminoacyl-tRNAs are limiting. Such a finding is important from the point of view of extending the SAVE approach to other kinds of viruses.

Translational Frameshifting

Another possible defect is that codon changes could promote translational frameshifting; that is, at some codon pairs, the ribosome could shift into a different reading frame, and then arrive at an in-frame stop codon after translating a spurious peptide sequence. This type frameshifting is an important regulatory event in some viruses. The present data reveal that all PV genomes carrying the AB mutant segment from residue 1513 to 2470 are non-viable. Furthermore, all genomes carrying this mutant region produce a novel protein band during in vitro translation of approximately 42-44 kDa (see FIG. 5A, marked by asterisk). This novel protein could be the result of a frameshift.

Examination of the sequence in the 1513-2470 interval reveals three potential candidate sites that conform to the slippery heptameric consensus sequence for -1 frameshifting in eukaryotes (X-XXY-YYZ) (Farabaugh, 1996). These sites are A-AAA-AAT at positions 1885 and 1948, and T-TTA-TTT at position 2119. They are followed by stop codons in the -1 frame at 1929, 1986 or 2149, respectively. The former two seem the more likely candidates to produce a band of the observed size.

To determine whether frameshifting is occurring, each of the three candidate regions is separately mutated so that it becomes unfavorable for frameshifting. Further, each of the

candidate stop codons is separately mutated to a sense codon. These six new point mutants are tested by *in vitro* translation. Loss of the novel 42-44 kDa protein upon mutation of the frameshifting site to an unfavorable sequence, and an increase in molecular weight of that protein band upon elimination of the stop codon, indicate that frameshifting is occurring. If frameshifting is the cause of the aberrant translation product, the viability of the new mutant that lacks the frameshift site is tested in the context of the 1513-2470 mutant segment. Clearly such a finding would be of significance for future genome designs, and if necessary, a frameshift filter may be incorporated in the software algorithm to avoid potential frameshift sites.

More detailed investigations of translational defects are conducted using various techniques including, but not limited to, polysome profiling, toeprinting, and luciferase assays of fusion proteins.

Polysome Profiling

Polysome profiling is a traditional method of examining translation. It is not high-throughput, but it is very well developed and understood. For polysome profiling, cell extracts are made in a way that arrests translation (with cycloheximide) and yet preserves the set of ribosomes that are in the act of translating their respective mRNAs (the “polysomes”). These polysomes are fractionated on a sucrose gradient, whereby messages associated with a larger number of ribosomes sediment towards the bottom. After fractionation of the gradient and analysis of RNA content using UV absorption, a polysome profile is seen where succeeding peaks of absorption correspond to mRNAs with N+1 ribosomes; typically 10 to 15 distinct peaks (representing the 40S ribosomal subunit, the 60S subunit, and 1, 2, 3, . . . 12, 13 ribosomes on a single mRNA) can be discerned before the peaks smudge together. The various fractions from the sucrose gradient are then run on a gel, blotted to a membrane, and analyzed by Northern analysis for particular mRNAs. This then shows whether that particular mRNA is primarily engaged with, say, 10 to 15 ribosomes (well translated), or 1 to 4 ribosomes (poorly translated).

In this study, for example, the wt virus, the PV-AB (codon deoptimized) virus, and its derivatives PV-AB⁷⁵⁵⁻¹⁵¹³, and PV-AB²⁹⁵⁴⁻³³⁸⁶, which have primarily N-terminal or C-terminal deoptimized segments, respectively, are compared. The comparison between the N-terminal and C-terminal mutant segments is particularly revealing. If codon deoptimization causes translation to be slow, or paused, then the N-terminal mutant RNA is associated with relatively few ribosomes (because the ribosomes move very slowly through the N-terminal region, preventing other ribosomes from loading, then zip through the rest of the message after traversing the deoptimized region). In contrast, the C-terminal mutant RNA are associated with a relatively large number of ribosomes, because many ribosomes are able to load, but because they are hindered near the C-terminus, they cannot get off the transcript, and the number of associated ribosomes is high.

Polysome analysis indicates how many ribosomes are actively associated with different kinds of mutant RNAs, and can, for instance, distinguish models where translation is slow from models where the ribosome actually falls off the RNA prematurely. Other kinds of models can also be tested.

Toeprinting

Toeprinting is a technique for identifying positions on an mRNA where ribosomes are slow or paused. As in polysome profiling, actively translating mRNAs are obtained, with their ribosomes frozen with cycloheximide but still associated; the mRNAs are often obtained from an *in vitro*

81

translation reaction. A DNA oligonucleotide primer complementary to some relatively 3' portion of the mRNA is used, and then extended by reverse transcriptase. The reverse transcriptase extends until it collides with a ribosome. Thus, a population of translating mRNA molecules generates a population of DNA fragments extending from the site of the primer to the nearest ribosome. If there is a site or region where ribosomes tend to pause (say, 200 bases from the primer), then this site or region will give a disproportionate number of DNA fragments (in this case, fragments 200 bases long). This then shows up as a "toeprint" (a band, or dark area) on a high resolution gel. This is a standard method for mapping ribosome pause sites (to within a few nucleotides) on mRNAs.

Chimeras with segments of deoptimized codons or codon pairs, wherein in different chimeras the segments are shifted slightly 5' or 3', are analyzed. If the deoptimized segments cause ribosomes to slow or pause, the toeprint shifts 5' or 3' to match the position of the deoptimized segment. Controls include wt viral RNA and several (harmlessly) shuffled viral RNAs. Controls also include pure mutant viral RNA (i.e., not engaged in translation) to rule out ribosome-independent effects of the novel sequence on reverse transcription.

The toeprint assay has at least two advantages. First, it can provide direct evidence for a paused ribosome. Second, it has resolution of a few nucleotides, so it can identify exactly which deoptimized codons or deoptimized codon pairs are causing the pause. That is, it may be that only a few of the deoptimized codons or codon pairs are responsible for most of the effect, and toe-printing can reveal that.

Dual Luciferase Reporter Assays of Fusion Proteins

The above experiments may suggest that certain codons or codon pairs are particularly detrimental for translation. As a high-throughput way to analyze effects of particular codons and codon pairs on translation, small test peptides are designed and fused to the N-terminus of sea pansy luciferase. Luciferase activity is then measured as an assay of the translatability of the peptide. That is, if the N-terminal peptide is translated poorly, little luciferase will be produced.

A series of eight 25-mer peptides are designed based on the experiments above. Each of the eight 25-mers is encoded 12 different ways, using various permutations of rare codons and/or rare codon pairs of interest. Using assembly PCR, these 96 constructs (8 25-mers×12 encodings) are fused to the N-terminus of firefly luciferase (F-luc) in a dicistronic, dual luciferase vector described above (see Example 5 and FIG. 6). A dual luciferase system uses both the firefly luciferase (F-Luc) and the sea pansy (*Renilla*) luciferase (R-Luc); these emit light under different biochemical conditions, and so can be separately assayed from a single tube or well. A dicistronic reporter is expressed as a single mRNA, but the control luciferase (R-Luc) is translated from one internal ribosome entry site (IRES), while the experimental luciferase (F-luc) (which has the test peptides fused to its N-terminus) is independently translated from its own IRES. Thus, the ratio of F-Luc activity to R-Luc activity is an indication of the translatability of the test peptide. See FIG. 6.

The resulting 96 dicistronic reporter constructs are transfected directly from the PCR reactions into 96 wells of HEK293 or HeLa cells. The firefly luciferase of the upstream cistron serves as an internal transfection control. Codon- or codon-pair-dependent expression of the sea pansy luciferase in the second cistron can be accurately determined as the ratio between R-Luc and F-Luc. This assay is high-

82

throughput in nature, and hundreds or even thousands of test sequences can be assayed, as necessary.

EXAMPLE 13

Design and Synthesis of Attenuated Viruses Using Novel Alternative-Codon Strategy

The SAVE approach to re-engineering viruses for vaccine production depends on large-scale synonymous codon substitution to reduce translation of viral proteins. This can be achieved by appropriately modulating the codon and codon pair bias, as well as other parameters such as RNA secondary structure and CpG content. Of the four de novo PV designs, two (the shuffled codon virus, PV-SD, and the favored codon pair virus, PV-Max) resulted in little phenotypic change over the wt virus. The other two de novo designs (PV-AB and PV-Min) succeeded in killing the virus employing only synonymous substitutions through two different mechanisms (drastic changes in codon bias and codon pair bias, respectively). The live-but-attenuated strains were constructed by subcloning regions from the inactivated virus strains into the wt.

A better understanding of the underlying mechanisms of viral attenuation employing large scale synonymous substitutions facilitates predictions of the phenotype and expression level of a synthetic virus. Ongoing studies address questions relating to the effect of the total number of alterations or the density of alterations on translation efficiency; the effect of the position of dense regions on translation; the interaction of codon and codon pair bias; and the effect of engineering large numbers of short-range RNA secondary structures into the genome. It is likely that there is a continuum between the wt and inactivated strains, and that any desired attenuation level can be engineered into a weakened strain. However, there may be hard limits on the attenuation level that can be achieved for any infection to be at self-sustaining and hence detectable. The 15⁴⁴² encodings of PV proteins constitutes a huge sequence space to explore, and various approaches are being utilized to explore this sequence space more systematically. These approaches include, first, developing a software platform to help design novel attenuated viruses, and second, using this software to design, and then synthesize and characterize, numerous new viruses that explore more of the sequence space, and answer specific questions about how alternative encodings cause attenuation. Additionally, an important issue to consider is whether dangerous viruses might accidentally be created by apparently harmless shuffling of synonymous codons.

Development of Software for Computer-Based Design of Viral Genomes and Data Analysis

Designing synthetic viruses requires substantial software support for (1) optimizing codon and codon-pair usage and monitoring RNA secondary structure while preserving, embedding, or removing sequence specific signals, and (2) partitioning the sequence into oligonucleotides that ensure accurate sequence-assembly. The prototype synthetic genome design software tools are being expanded into a full environment for synthetic genome design. In this expanded software, the gene editor is conceptually built around constraints instead of sequences. The gene designer works on the level of specifying characteristics of the desired gene (e.g., amino acid sequence, codon/codon-pair distribution, distribution of restriction sites, and RNA secondary structure constraints), and the gene editor algorithmically designs a DNA sequence realizing these constraints. There are many constraints, often interacting with each other, including, but not limited to, amino acid sequence, codon bias, codon pair

bias, CG dinucleotide content, RNA secondary structure, cis-acting nucleic acid signals such as the CRE, splice sites, polyadenylation sites, and restriction enzyme recognition sites. The gene designer recognizes the existence of these constraints, and designs genes with the desired features while automatically satisfying all constraints to a pre-specified level.

The synthesis algorithms previously developed for embedding/removing patterns, secondary structures, overlapping coding frames, and adhering to codon/codon-pair distributions are implemented as part of the editor, but more important are algorithms for realizing heterogeneous combinations of such preferences. Because such combinations lead to computationally intractable (NP-complete) problems, heuristic optimization necessarily plays an important role in the editor. Simulated annealing techniques are employed to realize such designs; this is particularly appropriate as simulated annealing achieved its first practical use in the early VLSI design tools.

The full-featured gene design programming environment is platform independent, running in Linux, Windows and MacOS. The system is designed to work with genomes on a bacterial or fungal (yeast) scale, and is validated through the synthesis and evaluation of the novel attenuated viral designs described below.

Virus Designs with Extreme Codon Bias in One or a Few Amino Acids

For a live vaccine, a virus should be as debilitated as possible, short of being inactivated, in which case there is no way to grow and manufacture the virus. One way of obtaining an optimally debilitated is to engineer the substitution of rare codons for just one or a few amino acids, and to create a corresponding cell line that overexpresses the rare tRNAs that bind to those rare codons. The virus is then able to grow efficiently in the special, permissive cell line, but is inviable in normal host cell lines. Virus is grown and manufactured using the permissive cell line, which is not only very convenient, but also safer than methods used currently used for producing live attenuated vaccines.

With the sequencing of the human genome, information regarding copy number of the various tRNA genes that read rare codons is available. Based on the literature (e.g., Lavner and Kotlar, 2005), the best rare codons for present purposes are CTA (Leu), a very rare codon which has just two copies of the cognate tRNA gene; TCG (Ser), a rare codon with four copies of the cognate tRNA gene; and CCG (Pro), a rare codon with four copies of the cognate tRNA gene (Lavner and Kotlar, 2005). The median number of copies for a tRNA gene of a particular type is 9, while the range is 2 to 33 copies (Lavner and Kotlar, 2005). Thus, the CTA codon is not just a rare codon, but is also the one codon with the fewest cognate tRNA genes. These codons are not read by any other tRNA; for instance, they are not read via wobble base pairing.

Changing all the codons throughout the virus genome coding for Leu (180 codons), Ser (153), and Pro (119) to the rare synonymous codons CTA, TCG, or CCG, respectively, is expected to create severely debilitated or even non-viable viruses. Helper cells that overexpress the corresponding rare tRNAs can then be created. The corresponding virus is absolutely dependent on growing only in this artificial culture system, providing the ultimate in safety for the generation of virus for vaccine production.

Four high-priority viruses are designed and synthesized: all Leu codons switched to CTA; all Ser codons switched to TCG; all Pro codons switched to CCG; and all Leu, Ser, and Pro codons switched to CTA, TCG, and CCG, respectively,

in a single virus. In one embodiment, these substitutions are made only in the capsid region of the virus, where a high rate of translation is most important. In another embodiment, the substitutions are made throughout the virus.

5 CG Dinucleotide Bias Viruses

With few exceptions, virus genomes under-represent the dinucleotide CpG, but not GpC (Karlin et al., 1994). This phenomenon is independent of the overall G+C content of the genome. CpG is usually methylated in the human genome, so that single-stranded DNA containing non-methylated CpG dinucleotides, as often present in bacteria and DNA viruses, are recognized as a pathogen signature by the Toll-like receptor 9. This leads to activation of the innate immune system. Although a similar system has not been shown to operate for RNA viruses, inspection of the PV genome suggests that PV has selected against synonymous codons containing CpG to an even greater extent than the significant under-representation of CpG dinucleotides in humans. This is particularly striking for arginine codons. Of the six synonymous Arg codons, the four CG containing codons (CGA, CGC CGG, CCU) together account for only 24 of all 96 Arg codons while the remaining two (AGA, AGG) account for 72. This in contrast to the average human codon usage, which would predict 65 CG containing codons and 31 AGA/AGO codons. In fact, two of the codons under-represented in PV are frequently used in human cells (CGC, CGG). There are two other hints that CG may be a disadvantageous dinucleotide in PV. First, in the codon pair-deoptimized virus, many of the introduced rare codon pairs contain CG as the central dinucleotide of the codon pair hexamer. Second, when Burns et al. (2006) passaged their codon bias-deoptimized virus and sequenced the genomes, it was observed that these viruses evolved to remove some CG dinucleotides.

Thus, in one high-priority redesigned virus, most or all Arg codons are changed to CGC or CGG (two frequent human codons). This does not negatively affect translation and allows an assessment of the effect of the CpG dinucleotide bias on virus growth. The increased C+G content of the resulting virus requires careful monitoring of secondary structure; that is, changes in Arg codons are not allowed to create pronounced secondary structures.

45 Modulating Codon-Bias and Codon-Pair Bias Simultaneously.

Codon bias and codon-pair bias could conceivably interact with each other at the translational level. Understand this interaction may enable predictably regulation of the translatability of any given protein, possibly over an extreme range.

50 If we represent wild type polio codon bias and codon pair bias as 0, and the worst possible codon bias and codon pair bias as -1, then four high-priority viruses are the (-0.3, -0.3), (-0.3, -0.6), (-0.6, -0.3), and (-0.6, -0.6) viruses. These viruses reveal how moderately poor or very poor codon bias interacts with moderately poor or very poor codon pair virus. These viruses are compared to the wild type and also to the extreme PV-AB (-1, 0) and PV-Min (0, -1) designs.

60 Modulating RNA Secondary Structure

The above synthetic designs guard against excessive secondary structures. Two additional designs systematically avoid secondary structures. These viruses are engineered to have wt codon and codon-pair bias with (1) provably minimal secondary structure, and (2) many small secondary structures sufficient to slow translation.

85

Additional Viral Designs

Additional viral designs include full-genome codon bias and codon-pair bias designs; non-CG codon pair bias designs; reduced density rare codon designs; and viruses with about 150 rare codons, either spread through the capsid region, or grouped at the N-terminal end of the capsid, or grouped at the C-terminal end of the capsid.

EXAMPLE 14

Testing the Potential for Accidentally Creating Viruses of Increased Virulence

It is theoretically possible that redesigning of viral genomes with the aim of attenuating these viruses could accidentally make a virus more virulent than the wt virus. Because protein sequences are not altered in the SAVE procedure, this outcome is unlikely. Nevertheless, it is desirable to experimentally demonstrate that the SAVE approach is benign.

Out of the possible 10^{442} sequences that could possibly encode PV proteins, some reasonably fit version of PV likely arose at some point in the past, and evolved to a local optimum (as opposed to a global optimum). The creation of a new version of PV with the same protein coding capacity but a very different set of codons places this new virus in a different location on the global fitness landscape, which could conceivably be close to a different local optimum than wt PV. Conceivably, this new local optimum could be better than the wild type local optimum. Thus, it is just barely possible that shuffling synonymous codons might create a fitter virus.

To investigate this possibility, 13 PV genomes are redesigned and synthesized: one virus with the best possible codon bias; one virus with the best possible codon pair bias (i.e., PV-Max); one virus with the best possible codon and codon pair bias; and 10 additional viruses with wt codon and codon pair bias, but shuffled synonymous codons. Other parameters, such as secondary structure, C+G content, and CG dinucleotide content are held as closely as possible to wt levels.

These 13 viruses may each be in a very different location of the global fitness landscape from each other and from the wild type. But none of them is at a local optimum because they have not been subject to selection. The 13 viruses and the wt are passaged, and samples viruses are taken at the 1st, 10th, 20th, and 50th passages. Their fitness is compared to each other and to wt by assessing plaque size, plaque-forming units/ml in one-step growth curves, and numbers of particles formed per cell. See Example 1. Five examples of each of the 13 viruses are sequenced after the 10th, 20th, and 50th passage. Select passage isolates are tested for pathogenicity in CD155tg mice, and LD₅₀'s are determined. These assays reveal whether any of the viruses are fitter than wt, and provide a quantitative measure of the risk of accidental production of especially virulent viruses. The 10 viruses with wt levels of codon and codon pair bias also provide information on the variability of the fitness landscape at the encoding level.

In view of the possibility that a fitter virus could emerge, and that a fitter virus may be a more dangerous virus, these experiments are conducted in a BSL3 laboratory. After the 10th passage, phenotypes and sequences are evaluated and the susceptibility of emerging viruses to neutralization by PV-specific antibodies is verified. The experiment is stopped and reconsidered if any evidence of evolution towards a significantly fitter virus, or of systematic evolution towards new protein sequences that evade antibody neutralization, is

86

obtained. Phenotypes and sequences are similarly evaluated after passage 20 before proceeding to passage 50. Because the synthetic viruses are created to encode exactly the same proteins as wt virus, the scope for increased virulence seems very limited, even if evolution towards (slightly) increased fitness is observed.

EXAMPLE 15

Extension of SAVE Approach to Virus Systems Other than Poliovirus

Notwithstanding the potential need for a new polio vaccine to combat the potential of reversion in the closing stages of the global effort at polio eradication, PV has been selected in the present studies primarily as a model system for developing SAVE. SAVE has, however, been developed with the expectation that this approach can be extended to other viruses where vaccines are needed. This extension of the SAVE strategy is herein exemplified by application to Rhinovirus, the causative agent of the common cold, and to influenza virus.

Adaptation of SAVE to Human Rhinovirus—a Virus Closely Related to Poliovirus

Two model rhinoviruses, HRV2 and HRV14, were selected to test the SAVE approach for several reasons: (1) HRV2 and HRV14 represent two members of the two different genetic subgroups, A and B (Ledford et al., 2004); (2) these two model viruses use different receptors, LDL-receptor and ICAM-1, respectively (Greve et al., 1989; Hofer et al., 1994); both viruses as well as their infectious cDNA clones have been used extensively, and most applicable materials and methods have been established (Altmeyer et al., 1991; Gerber et al., 2001); and (4) much of the available molecular knowledge of rhinoviruses stems from studies of these two serotypes.

The most promising SAVE strategies developed through the PV experiments are applied to the genomes of HRV2 and HRV14. For example, codons, codon pairs, secondary structures, or combinations thereof, are deoptimized. Two to three genomes with varying degrees of attenuation are synthesized for each genotype. Care is taken not to alter the CRE, a critical RNA secondary structure of about 60 nucleotides (Gerber et al., 2001; Goodfellow et al., 2000; McKnight, 2003). This element is vital to the replication of picornaviruses and thus the structure itself must be maintained when redesigning genomes. The location of the CRE within the genome varies for different picornaviruses, but is known for most families (Gerber et al., 2001; Goodfellow et al., 2000; McKnight, 2003), and can be deduced by homology modeling for others where experimental evidence is lacking. In the case of HRV2 the CRE is located in the RNA sequence corresponding to the nonstructural protein 2A^{Pro}; and the CRE of HRV14 is located in the VP1 capsid protein region (Gerber et al., 2001; McKnight, 2003).

The reverse genetics system to derive rhinoviruses from DNA genome equivalents is essentially the same as described above for PV, with the exception that transfections are done in HeLa-H1 cells at 34° C. in Hepes-buffered culture medium containing 3 mM Mg⁺⁺ to stabilize the viral capsid. The resulting synthetic viruses are assayed in tissue culture to determine the PFU/IU ratio. See Example 3. Plaque size and kinetics in one-step growth curves are also assayed as described. See Example 2. Because the SAVE process can be applied relatively cheaply to all 100 or so relevant rhinoviruses, it is feasible to produce a safe and effective vaccine for the common cold using this approach.

Adaptation of SAVE to Influenza A Virus—a Virus Unrelated to Poliovirus

The most promising SAVE design criteria identified from PV experimentation are used to synthesize codon-deoptimized versions of influenza virus. The influenza virus is a “segmented” virus consisting of eight separate segments of RNA; each of these can be codon-modified. The well established ambisense plasmid reverse genetics system is used for generating variants of influenza virus strain A/PR/8/34. This eight-plasmid system is a variation of what has been described previously (Hoffmann et al., 2000), and has been kindly provided by Drs. P. Palese and A. Garcia-Sastre. Briefly, the eight genome segments of influenza each contained in a separate plasmid are flanked by a Pol I promoter at the 3' end and Pol I terminator at the 5' end on the antisense strand. This cassette in turn is flanked by a cytomegalovirus promoter (a Pol II promoter) at the 5' end and a polyadenylation signal at the 3' end on the forward strand (Hoffmann et al., 2000). Upon co-transfection into co-cultured 293T and MDCK cells, each ambisense expression cassette produces two kinds of RNA molecules. The Pol II transcription units on the forward strand produce all influenza mRNAs necessary for protein synthesis of viral proteins. The Pol I transcription unit on the reverse strand produces (-) sense genome RNA segments necessary for assembly of ribonucleoprotein complexes and encapsidation. Thus, infectious influenza A/PR/8/34 particles are formed (FIG. 10). This particular strain of the H1 N1 serotype is relatively benign to humans. It has been adapted for growth in tissue culture cells and is particularly useful for studying pathogenesis, as it is pathogenic in BALB/c mice.

When synthesizing segments that are alternatively spliced (NS and M), care is taken not to destroy splice sites and the alternative reading frames. In all cases the terminal 120 nt at either end of each segment are excluded, as these sequences are known to contain signals for RNA replication and virus assembly. At least two versions of each fragment are synthesized (moderate and maximal deoptimization). Viruses in which only one segment is modified are generated, the effect is assessed, and more modified segments are introduced as needed. This is easy in this system, since each segment is on a separate plasmid.

Virus infectivity is titered by plaque assay on MDCK cells in the presence of 1 ug/ml (TPCK)-trypsin. Alternatively, depending on the number of different virus constructs, a 96-well ELISA is used to determine the titer of various viruses as cell infectious units on MDCK cells essentially as described above for PV. See Example 3. The only difference

is that now a HA-specific antibody is used to stain infected cells. In addition, the relative concentration of virions are determined via hemagglutination (HA) assay using chicken red blood cells (RBC) (Charles River Laboratories) using standard protocols (Kendal et al., 1982). Briefly, virus suspensions are 2-fold serially diluted in PBS in a V-bottom 96 well plates. PBS alone is used as an assay control. A standardized amount of RBCs is added to each well, and the plates are briefly agitated and incubated at room temperature for 30 minutes. HA titers are read as the reciprocal dilution of virus in the last well with complete hemagglutination. While HA-titer is a direct corollary of the amount of particles present, PFU-titer is a functional measure of infectivity. By determining both measures, a relative PFU/HA-unit ratio is calculated similar to the PFU/particle ratio described in the PV experiments. See Example 3. This addresses the question whether codon- and codon pair-deoptimized influenza viruses also display a lower PFU/particle as observed for PV.

Virulence Test

The lethal dose 50 (LD_{50}) of the parental NPIR/8/34 virus is first determined for mice and synthetic influenza viruses are chosen for infection of BALB/c mice by intranasal infection. Methods for determining LD_{50} values are well known to persons of ordinary skill in the art (see Reed and Muench, 1938, and Example 4). The ideal candidate viruses display a low infectivity (low PFU titer) with a high virion concentration (high HA-titer). Anesthetized mice are administered 25 μ l of virus solution in PBS to each nostril containing 10-fold serial dilutions between 10^2 to 10^7 PFU of virus. Mortality and morbidity (weight loss, reduced activity) are monitored twice daily for up to three weeks. LD_{50} is calculated by the method of Reed and Muench (1938). For the A/PR/8/34 wt virus the expected LD_{50} is around 10^3 PFU (Talon et al., 2000), but may vary depending on the particular laboratory conditions under which the virus is titered.

Adaptation of SAVE to Dengue, HIV, Rotavirus, and SARS

Several viruses were selected to further test the SAVE approach. Table 8 identifies the coding regions of each of Dengue, HIV, Rotavirus (two segments), and SARS, and provides nucleotide sequences for parent viruses and exemplary viral genome sequences having deoptimized codon pair bias. As described above, codon pair bias is determined for a coding sequence, even though only a portion (subsequence) may contain the deoptimizing mutations.

TABLE 8

Nucleotide sequence and codon pair bias of parent and codon pair bias-reduced coding regions						
Virus	Parent sequence			Codon pair bias-reduced sequence		
	SEQ ID NO:	CDS	CPB	SEQ ID NO:	deoptimized segment*	CPB*
Flu PB1	13	25-2298	0.0415	14	531-2143	-0.2582
Flu PB1-RR	"	"	"	15	531-1488	-0.1266
Flu PB2	16	28-2307	0.0054	17	33-2301	-0.3718
Flu PA	18	25-2175	0.0247	19	30-2171	-0.3814
Flu HA	20	33-1730	0.0184	21	180-1655	-0.3627
Flu NP	22	46-1542	0.0069	23	126-1425	-0.3737
Flu NA	24	21-1385	0.0037	25	123-1292	-0.3686
Flu M	26		0.0024			
Flu NS	27	27-719	-0.0036	28	128-479	-0.1864
Rhinovirus	29	619-7113	0.051	30		-0.367

TABLE 8-continued

Nucleotide sequence and codon pair bias of parent and codon pair bias-reduced coding regions					
Virus	Parent sequence		Codon pair bias-reduced sequence		
	SEQ ID NO:	CDS	CPB	SEQ ID NO:	deoptimized segment*
Rhinovirus 14	31	629-7168	0.046	32	-0.418
Dengue HIV	33 35	95-10273 336-1634 1841-4585 4644-5102 5858-7924 8343-8963	0.0314 0.0656	34 36	-0.4835 -0.3544
Rotavirus Seg.1	37	12-3284	0.0430	38	-0.2064
Rotavirus Seg.2	39	37-2691	0.0375	40	-0.2208
SARS	41	265-13398 13416-21485 21492-25259 26398-27063	0.0286	42	-0.4393

*CPB can be reduced by deoptimizing an internal segment smaller than the complete coding sequence. Nevertheless, CPB is calculated for the complete CDS.

EXAMPLE 16

Assessment of Poliovirus and Influenza Virus Vaccine Candidates in Mice

The ability of deoptimized viruses to vaccinate mice against polio or influenza is tested.

Poliovirus Immunizations, Antibody Titers, and Wt Challenge Experiments

The working hypothesis is that a good vaccine candidate combines a low infectivity titer with a high virion titer. This ensures that a high amount of virus particles (i.e., antigen) can be injected while at the same time having a low risk profile. Thus, groups of five CD155tg mice will be injected intraperitoneally with 10^3 , 10^4 , 10^5 , and 10^6 PFU of PV(Mahoney) (i.e., wild-type), PV1 Sabin vaccine strain, PV⁴⁸²⁴⁷⁰⁻²⁹⁵⁴, PV-Min⁷⁵⁵⁻²⁴⁷⁰, or other promising attenuated polioviruses developed during this study. For the wild-type, 1 PFU is about 100 viral particles, while for the attenuated viruses, 1 PFU is roughly 5,000 to 100,000 particles. Thus, injection with equal number of PFUs means that 50 to 1000-fold more particles of attenuated virus are being injected. For wt virus injected intraperitoneally, the LD₅₀ is about 10^6 PFU, or about 10^8 particles. Accordingly, some killing is expected with the highest doses but not with the lower doses.

Booster shots of the same dose are given one week after and four weeks after the initial inoculation. One week following the second booster, PV-neutralizing antibody titers are determined by plaque reduction assay. For this purpose, 100 PFU of wt PV(M) virus are incubated with 2-fold serial dilutions of sera from immunized mice. The residual number of PFU is determined by plaque assays. The neutralizing antibody titer is expressed as the reciprocal of the lowest serum dilution at which no plaques are observed.

Four weeks after the last booster, immunized mice and non-immunized controls are challenged with a lethal dose of PV(M) wt virus (10^6 PFU intraperitoneally; this equals 100 times LD₅₀, and survival is monitored.

Influenza Immunizations, Antibody Titers, and Wt Challenge Experiments

For vaccination experiments, groups of 5 BALB/c mice are injected with wt and attenuated influenza viruses intra-

peritoneally at a dose of 0.001, 0.01, 0.1, and 1.0 LD₅₀. Booster vaccinations are given at the same intervals described above for PV. Influenza antibody titers one week after the second booster are determined by an inhibition of hemagglutination (HI) assay following standard protocols (Kendal et al., 1982). Briefly, sera from immunized and control mice treated with receptor destroying enzyme (RDE; Sigma, St Louis, Mo.) are 2-fold serially diluted and mixed with 5 HA-units of A/PR/8/34 virus in V-bottom 96 wells. RBCs are then added and plates are processed as above for the standard HA-assay. Antibody titers are expressed as the reciprocal dilution that results in complete inhibition of hemagglutination.

Three weeks after the last booster vaccination, mice are challenged infra-nasally with 100 or 1000 LD₅₀ of A/PR/8/34 parental virus (approximately 10^5 and 10^6 PFU), and survival is monitored.

Animal Handling

Transgenic mice expressing the human poliovirus receptor CD155 (CD155tg) were obtained from Dr. Nomoto, The Tokyo University. The CD155tg mouse colony is maintained by the State University of New York (SUNY) animal facility. BALB/c mice are obtained from Taconic (Germantown, N.Y.). Anesthetized mice are inoculated using 25-gauge hypodermic needles with 30 μ l of viral suspension by intravenous, intraperitoneal or intracerebral route or 50 μ l by the intranasal route. Mice of both sexes between 6-24 weeks of age are used. Mice are the most economical model system for poliovirus and influenza virus research. In addition, in the case of PV, the CD155tg mouse line is the only animal model except for non-human primates. Mice also provide the safest animal model since no virus spread occurs between animals for both poliovirus and influenza virus.

All mice are housed in SUNY's state of the art animal facility under the auspices of the Department of Laboratory Animal Research (DLAR) and its veterinary staff. All animals are checked twice weekly by the veterinary staff. Virus-infected animals are checked twice daily by the investigators and daily by the veterinary staff. All infection experiments are carried out in specially designated maximum isolation rooms within the animal facility. After con-

clusion of an experiment, surviving mice are euthanized and cadavers are sterilized by autoclaving. No mouse leaves the virus room alive.

In the present study, mice are not subjected to any surgical procedure besides intravenous, intracerebral, intraperitoneal, intramuscular or intranasal inoculation, the injection of anesthetics, and the collection of blood samples. For vaccination experiments, blood samples are taken prior and after vaccination for detection of virus-specific antibodies. To this end, 50-100 μl are collected from mice the day before injection and one week following the second booster vaccination. A maximum of two blood samples on individual animals are collected at least four weeks apart. Animals are anesthetized and a sharp scalpel is used to cut off 2 mm of tail. Blood is collected with a capillary tube. Subsequent sampling is obtained by removing scab on the tail. If the tail is healed, a new 2-mm snip of tail is repeated.

All animal experiments are carried out following protocols approved by the SUNY Institutional Animal Care and Use Committee (IACUC). Euthanasia is performed by trained personnel in a CO_2 gas chamber according to the recommendation of the American Veterinary Medical Association. Infection experiments are conducted under the latest the ABSL 2/polio recommendations issued by the Centers for Disease Control and Prevention (CDC).

EXAMPLE 17

Codon Pair Bias Algorhythm—Codon Pair Bias and Score Matrix

In most organisms, there exists a distinct codon bias, which describes the preferences of amino acids being encoded by particular codons more often than others. It is widely believed that codon bias is connected to protein translation rates. In addition, each species has specific preferences as to whether a given pair of codons appear as neighbors in gene sequences, something that is called codon-pair bias.

To quantify codon pair bias, we define a codon pair distance as the log ratio of the observed over the expected number of occurrences (frequency) of codon pairs in the genes of an organism. Although the calculation of the observed frequency of codon pairs in a set of genes is straightforward, the expected frequency of a codon pair is calculated as in Gutman and Hatfield, Proc. Natl. Acad. Sci. USA, 86:3699-3703, 1989, and is independent of amino acid and codon bias. To achieve that, the expected frequency is calculated based on the relative proportion of the number of times an amino acid is encoded by a specific codon. In short:

$$\text{codon pair score} = \log \left(\frac{F(AB)}{\frac{F(A) \times F(B)}{F(X) \times F(Y)} \times F(XY)} \right),$$

where the codon pair AB encodes for amino acid pair XY and F denotes frequency (number of occurrences).

In this scheme we can define a 64×64 codon-pair distance matrix with all the pairwise costs as defined above. Any m-residue protein can be rated as using over- or under-represented codon pairs by the average of the codon pair scores that comprise its encoding.

Optimization of a Gene Encoding Based on Codon Pair Bias

To examine the effects of codon pair bias on the translation of specific proteins, we decided to change the codon

pairs while keeping the same codon distribution. So we define the following problem: Given an amino acid sequence and a set of codon frequencies (codon distribution), change the DNA encoding of the sequence such that the codon pair score is optimized (usually minimized or maximized).

Our problem, as defined above, can be associated with the Traveling Salesman Problem (TSP). The traveling salesman problem is the most notorious NP-complete problem. This is a function of its general usefulness, and because it is easy to explain to the public at large. Imagine a traveling salesman who has to visit each of a given set of cities by car. What is the shortest route that will enable him to do so and return home, thus minimizing his total driving?

TSP Heuristics

Almost any flavor of TSP is going to be NP-complete, so the right way to proceed is with heuristics. These are often quite successful, typically coming within a few percent of the optimal solution, which is close enough for most applications and in particular for our optimized encoding.

Minimum spanning trees—A simple and popular heuristic, especially when the sites represent points in the plane, is based on the minimum spanning tree of the points. By doing a depth-first search of this tree, we walk over each edge of the tree exactly twice, once going down when we discover the new vertex and once going up when we backtrack. We can then define a tour of the vertices according to the order in which they were discovered and use the shortest path between each neighboring pair of vertices in this order to connect them. This path must be a single edge if the graph is complete and obeys the triangle inequality, as with points in the plane. The resulting tour is always at most twice the length of the minimum TSP tour. In practice, it is usually better, typically 15% to 20% over optimal. Further, the time of the algorithm is bounded by that of computing the minimum spanning tree, only $O(n \lg n)$ in the case of points in the plane.

Incremental insertion methods—A different class of heuristics inserts new points into a partial tour one at a time (starting from a single vertex) until the tour is complete. The version of this heuristic that seems to work best is furthest point insertion: of all remaining points, insert the point v into partial tour T such that

$$\max_{v \in V} \min_{i=1}^{|T|} (d(v, v_i) + d(v, v_{i+1})).$$

The minimum ensures that we insert the vertex in the position that adds the smallest amount of distance to the tour, while the maximum ensures that we pick the worst such vertex first. This seems to work well because it first “roughs out” a partial tour before filling in details. Typically, such tours are only 5% to 10% longer than optimal.

k-optimal tours—Substantially more powerful are the Kernighan-Lin, or k-opt class of heuristics. Starting from an arbitrary tour, the method applies local refinements to the tour in the hopes of improving it. In particular, subsets of $k \geq 2$ edges are deleted from the tour and the k remaining subchains rewired in a different way to see if the resulting tour is an improvement. A tour is k-optimal when no subset of k edges can be deleted and rewired so as to reduce the cost of the tour. Extensive experiments suggest that 3-optimal tours are usually within a few percent of the cost of optimal tours. For $k > 3$, the computation time increases considerably faster than solution quality. Two-opting a tour is a fast and effective way to improve any other heuristic. Simulated

annealing provides an alternate mechanism to employ edge flips to improve heuristic tours.

Algorithm for Solving the Optimum Encoding Problem

Our problem as defined is associated with the problem of finding a traveling salesman path (not tour) under a 64-country metric. In this formulation, each of the 64 possible codons is analogous to a country, and the codon multiplicity modeled as the number of cities in the country. The codon-pair bias measure is reflected as the country distance matrix.

The real biological problem of the design of genes encoding specific proteins using a given set of codon multiplicities so as to optimize the gene/DNA sequence under a codon-pair bias measure is slightly different. What is missing in our model in the country TSP model is the need to encode specific protein sequences. The DNA triplet code partitions the 64 codons into 21 equivalence classes (coding for each of the 20 possible amino acids and a stop symbol). Any given protein/amino acid sequence can be specified by picking an arbitrary representative of the associated codon equivalence class to encode it.

Because of the special restrictions and the nature of our problem, as well as its adaptability to application of additional criteria in the optimization, we selected the Simulated annealing heuristic to optimize sequences. The technique is summarized below.

Simulated Annealing Heuristic

Simulated annealing is a heuristic search procedure that allows occasional transitions leading to more expensive (and hence inferior) solutions. This may not sound like a win, but it serves to help keep our search from getting stuck in local optima.

The inspiration for simulated annealing comes from the physical process of cooling molten materials down to the solid state. In thermodynamic theory, the energy state of a system is described by the energy state of each of the particles constituting it. The energy state of each particle jumps about randomly, with such transitions governed by the temperature of the system. In particular, the probability $P(e_i, e_j, T)$ of transition from energy e_i to e_j at temperature T is given by:

$$P(e_i, e_j, T) = e^{(e_j - e_i)/(k_B T)}$$

where k_B is a constant, called Boltzmann's constant. What does this formula mean? Consider the value of the exponent under different conditions. The probability of moving from a high-energy state to a lower-energy state is very high. However, there is also a nonzero probability of accepting a transition into a high-energy state, with small energy jumps much more likely than big ones. The higher the temperature, the more likely such energy jumps will occur.

What relevance does this have for combinatorial optimization? A physical system, as it cools, seeks to go to a minimum-energy state. For any discrete set of particles, minimizing the total energy is a combinatorial optimization problem. Through random transitions generated according to the above probability distribution, we can simulate the physics to solve arbitrary combinatorial optimization problems.

As with local search, the problem representation includes both a representation of the solution space and an appropriate and easily computable cost function $C(s)$ measuring the quality of a given solution. The new component is the cooling schedule, whose parameters govern how likely we are to accept a bad transition as a function of time.

At the beginning of the search, we are eager to use randomness to explore the search space widely, so the probability of accepting a negative transition should be high.

As the search progresses, we seek to limit transitions to local improvements and optimizations. The cooling schedule can be regulated by the following parameters:

Initial system temperature—Typically $t_1=1$.
Temperature decrement function—Typically $t_k=a \cdot t_{k-1}$, where $0.8 \leq a \leq 0.99$. This implies an exponential decay in the temperature, as opposed to a linear decay.

Number of iterations between temperature change—Typically, 100 to 1,000 iterations might be permitted before lowering the temperature.

Acceptance criteria—A typical criterion is to accept any transition from s_i to s_i+1 when $C(s_i+1) < C(s_i)$ and to accept a negative transition whenever

$$e^{-\frac{(C(s_i) - C(s_{i+1}))}{c \cdot t_i}} \geq r,$$

where r is a random number $0 \leq r < 1$. The constant c normalizes this cost function, so that almost all transitions are accepted at the starting temperature.

Stop criteria—Typically, when the value of the current solution has not changed or improved within the last iteration or so, the search is terminated and the current solution reported.

In expert hands, the best problem-specific heuristics for TSP can slightly outperform simulated annealing, but the simulated annealing solution works easily and admirably.

REFERENCES

- Alexander, H. E., G. Koch, I. M. Mountain, K. Sprunt, and O. Van Damme. 1958. Infectivity of ribonucleic acid of poliovirus on HeLa cell mono-layers Virology. 5:172-3.
- Altmeyer, R., A. D. Murdin, J. J. Harber, and E. Wimmer. 1991. Construction and characterization of poliovirus/rhinovirus antigenic hybrid. Virology. 184:636-44.
- Ansardi, D. C., D. C. Porter, and C. D. Morrow. 1993. Complementation of a poliovirus defective genome by a recombinant vaccinia virus which provides poliovirus P1 capsid precursor in trans. J. Virol. 67:3684-3690.
- Belov, G. A., L. I. Romanova, E. A. Tolskaya, M. S. Kolesnikova, Y. A. Lazebnik, and V. I. Agol. 2003. The major apoptotic pathway activated and suppressed by poliovirus. J. Virol. 77:45-56.
- Buchan, J. R., L. S. Aucott, and I. Stansfield. 2006. tRNA properties help shape codon pair preferences in open reading frames. Nucl. Acids Res. 34:1015-27.
- Burns, C. C., J. Shaw, R. Campagnoli, J. Jorba, A. Vincent, J. Quay, and O. Kew. 2006. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. J. Virol. 80:3259-72.
- Cao, X., R. J. Kuhn, and E. Wimmer. 1993. Replication of poliovirus RNA containing two VPg coding sequences leads to a specific deletion event. J. Virol. 67:5572-5578.
- Carlini, D. B., and W. Stephan. 2003. In vivo introduction of unpreferred synonymous codons into the *Drosophila* Adh gene results in reduced levels of ADH protein. Genetics 163:239-243.
- Cello, J., A. V. Paul, and E. Wimmer. 2002. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template Science. 297: 1016-1018.

- Cheng, L., and E. Goldman. 2001. Absence of effect of varying Thr-Leu codon pairs on protein synthesis in a T7 system. *Biochemistry*. 40:6102-6.
- Cohen, B., and S. Skiena. 2003. Natural selection and algorithmic design of mRNA. *J. Comput Biol.* 10:419-432.
- Coligan, J., A. Kruisbeek, D. Margulies, E. Shevach, and W. Strober, eds. (1994) *Current Protocols in Immunology*, Wiley & Sons, Inc., New York.
- Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* 16:10881-90.
- Cram, P., S. G. Blitz, A. Monte, and A. M. Fendrick. 2001. Influenza. Cost of illness and consideration in the economic evaluation of new and emerging therapies. *Pharmacoeconomics*. 19:223-30.
- Crotty, S., C. E. Cameron, and R. Andino. 2001. RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc. Natl. Acad. Sci. U.S.A.* 98:6895-6900.
- Curran, J. F., E. S. Poole, W. P. Tate, and B. L. Gross. 1995. Selection of aminoacyl-tRNAs at sense codons: the size of the tRNA variable loop determines whether the immediate 3' nucleotide to the codon has a context effect. *Nucl. Acids Res.* 23:4104-8.
- Doma, M. K., and R. Parker. 2006. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*. 440:561-4.
- Dove, A. W., and V. R. Racaniello. 1997. Cold-adapted poliovirus mutants bypass a postentry replication block. *J. Virol.* 71:4728-4735.
- Enami, M., W. Luytjes, M. Krystal, and P. Palese. 1990. Introduction of site-specific mutations into the genome of influenza virus. *Proc. Natl. Acad. Sci. U.S.A.* 87:3802-5.
- Farabaugh, P. J. 1996. Programmed translational frame-shifting. *Microbiol Rev.* 60:103-34.
- Fedorov, A., S. Saxonov, and W. Gilbert. 2002. Regularities of context-dependent codon bias in eukaryotic genes. *Nucl. Acids Res.* 30:1192-7.
- Fodor, E., L. Devenish, O. G. Engelhardt, P. Palese, G. G. Brownlee, and A. Garcia-Sastre. 1999. Rescue of influenza A virus from recombinant DNA. *J. Virol.* 73:9679-82.
- Gabow, H. 1973. Ph.D. thesis. Stanford University, Stanford, Calif.
- Garcia-Sastre, A., and P. Palese. 1993. Genetic manipulation of negative-strand RNA virus genomes. *Annu. Rev. Microbiol.* 47:765-90.
- Georgescu, M. M., J. Balanant, A. Macadam, D. Otelea, M. Combieescu, A. A. Combieescu, R. Crainic, and F. Delpeyroux. 1997. Evolution of the Sabin type 1 poliovirus in humans: characterization of strains isolated from patients with vaccine-associated paralytic poliomyelitis. *J. Virol.* 71:7758-68.
- Gerber, K., E. Wimmer, and A. V. Paul. 2001. Biochemical and genetic studies of the initiation of human rhinovirus 2 RNA replication: identification of a cis-replicating element in the coding sequence of 2A(pro). *J. Virol.* 75:10979-10990.
- Girard, S., T. Couderc, J. Destombes, D. Thiesson, F. Delpeyroux, and B. Blondel. 1999. Poliovirus induces apoptosis in the mouse central nervous system. *J. Virol.* 73:6066-6072.
- Goodfellow, I., Y. Chaudhry, A. Richardson, J. Meredith, J. W. Almond, W. Barclay, and D. J. Evans. 2000. Identification of a cis-acting replication element within the poliovirus coding region. *J. Virol.* 74:4590-600.

- Greve, J. M., G. Davis, A. M. Meyer, C. P. Forte, S. C. Yost, C. W. Marlor, M. E. Kamarck, and A. McClelland. 1989. The major human rhinovirus receptor is ICAM-1. *Cell*. 56:839-47.
- Gustafsson, C., S. Govindarajan, and J. Minshull. 2004. Codon bias and heterologous protein expression. *Trends Biotechnol.* 22:346-353.
- Gutman, G. A., and G. W. Hatfield. 1989. Nonrandom utilization of codon pairs in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 86:3699-703.
- He, Y., V. D. Bowman, S. Mueller, C. M. Bator, J. Bella, X. Peng, T. S. Baker, E. Wimmer, R. J. Kuhn, and M. G. Rossmann. 2000. Interaction of the poliovirus receptor with poliovirus. *Proc. Natl. Acad. Sci. USA* 97:79-84.
- Hendley, J. O. 1999. Clinical virology of rhinoviruses *Adv Virus Res.* 54:453-66.
- Herold, J., and R. Andino. 2000. Poliovirus requires a precise 5' end for efficient positive-strand RNA synthesis. *J. Virol.* 74:6394-400.
- Hoekema, A., R. A. Kastlein, M. Vasser, and H. A. de Boer. 1987. Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. *Mol. Cell. Biol.* 7:2914-2924.
- Hofer, F., M. Gruenberger, H. Kowalski, H. Machat, M. Huettinger, E. Kuechler, and D. Blaas. 1994 Members of the low density lipoprotein receptor family mediate cell entry of a minor-group common cold virus. *Proc. Natl. Acad. Sci. U.S.A.* 91:1839-42.
- Hoffmann, E., G. Neumann, Y. Kawaoka, G. Hobom, and R. G. Webster. 2000. A DNA transfection system for generation of influenza: A virus from eight plasmids. *Proc. Natl. Acad. Sci. U.S.A.* 97:6108-13.
- Hogle, J. M. 2002. Poliovirus cell entry: common structural themes in viral cell entry pathways. *Annu. Rev. Microbiol.* 56:677-702.
- Holland, J. J., E. Domingo, J. C. de la Torre, and D. A. Steinhauer. 1990. Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. *J. Virol.* 64:3960-3962.
- Hsiao, L. L., F. Dangond, T. Yoshida, R. Hong, R. V. Jensen et al. 2001. A compendium of gene expression in normal human tissues. *Physiol. Genomics* 7:97-104.
- Irwin, B., J. D. Heck, and G. W. Hatfield. 1995. Codon pair utilization biases influence translational elongation step times. *J. Biol Chem.* 270:22801-6.
- Jang, S. K., M. V. Davies, R. J. Kaufman, and E. Wimmer. 1989. Initiation of protein synthesis by internal entry of ribosomes into the 5' nontranslated region of encephalomyocarditis virus RNA in vitro. *J. Virol.* 63:1651-1660.
- Jayaram, S., R. Reid, and D. V. Santi. 2005. GeMS: an advanced software package for designing synthetic genes. *Nucl. Acids Res.* 33:3011-3016.
- Johansen, L. K., and C. D. Morrow. 2000. The RNA encompassing the internal ribosome entry site in the poliovirus 5' nontranslated region enhances the encapsidation of genomic RNA. *Virology* 273:391-399.
- Joklik, W., and J. Darnell. 1961. The adsorption and early fate of purified poliovirus in HeLa cells. *Virology* 13:439-447.
- Kamps, B. S., C. Hoffmann, and W. Preiser (eds.) 2006. *Influenza Report*, 2006. Flying Publisher.

- Kaplan, G., and V. R. Racaniello. 1988. Construction and characterization of poliovirus subgenomic replicons. *J. Virol.* 62:1687-96.
- Karlin, S., W. Doerfler, and L. R. Cardon. 1994. Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *J. Virol.* 68:2889-97.
- Kendal, A. P., J. J. Skehel, and M. S. Pereira (eds.) 1982 Concepts and procedures for laboratory-based influenza surveillance. World Health Organization Collaborating Centers for Reference and Research on Influenza, Geneva.
- Kew, O., V. Morris-Glasgow, M. Landaverde, C. Burns, J. Shaw, Z. Garib, J. Andre, E. Blackman, C. J. Freeman, J. Jorba, R. Sutter, G. Tambini, L. Venczel, C. Pedreira, F. Laender, H. Shimizu, T. Yoneyama, T. Miyamura, H. van Der Avoort, M. S. Oberste, D. Kilpatrick, S. Cochi, M. Pallansch, and C. de Quadros. 2002. Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. *Science.* 296:356-9.
- Kilbourne, E. D. 2006. Influenza pandemics of the 20th century. *Emerg. Infect. Dis.* 12:9-14.
- Kitamura, N., B. L. Semler, P. G. Rothberg, G. R. Larsen, C. J. Adler, A. J. Dorner, E. A. Emini, R. Hanecak, J. Lee, S. van der Well, C. W. Anderson, and E. Wimmer. 1981. Primary structure, gene organization and poly-peptide expression of poliovirus RNA. *Nature.* 291: 547-553.
- Koike, S., C. Taya, T. Kurata, S. Abe, I. Ise, H. Yonekawa, and A. Nomoto. 1991. Transgenic mice susceptible to poliovirus. *Proc. Natl. Acad. Sci. U.S.A.* 88:951-955.
- Landsteiner, K. and E. Popper. 1909. Übertragung der Poliomyelitis acuta auf Affen. *Z. ImmunittatsForsch Orig.* 2:377-90.
- Lavner, Y., and D. Kotlar. 2005. Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene.* 345:127-38.
- Ledford, R. M., N. R. Patel, T. M. Demenczuk, A. Watanyar, T. Herblitz, M. S. Collett, and D. C. Pevear. 2004. VP1 sequencing of all human rhinovirus serotypes: insights into genus phylogeny and susceptibility to antiviral capsid-binding compounds. *J. Virol.* 78:3663-74.
- Luytjes, W., M. Krystal, M. Enami, J. D. Pavin, and P. Palese. 1989. Amplification, expression, and packaging of foreign gene by influenza virus. *Cell.* 59:1107-13.
- McKnight, K. L. 2003. The human rhinovirus internal cis-acting replication element (cre) exhibits disparate properties among serotypes. *Arch. Virol.* 148:2397-418.
- Molla, A., A. V. Paul, and E. Wimmer. 1991. Cell-free, de novo synthesis of poliovirus. *Science* 254:1647-1651.
- Mueller, S., D. Papamichail, J. R. Coleman, S. Skiena, and E. Wimmer. 2006. Reduction of the Rate of Poliovirus Protein Synthesis through Large-Scale Codon Deoptimization Causes Attenuation of Viral Virulence by lowering specific infectivity. *J. Virol.* 80:9687-9696.
- Mueller, S., E. Wimmer, and J. Cello. 2005. Poliovirus and poliomyelitis: a tale of guts, brains, and an accidental event. *Virus Res.* 111:175-193.
- Murdin, A. D., and E. Wimmer. 1989. Construction of a poliovirus type 1/type 2 antigenic hybrid by manipulation of neutralization antigenic site II. *J. Virol.* 63:5251-5257.
- Neumann, G., T. Watanabe, H. Ito, S. Watanabe, H. Goto, P. Gao, M. Hughes, D. R. Perez, R. Donis E. Hoffmann, G. Hobom, and Y. Kawaoka. 1999. Generation of

- influenza A viruses entirely from clone cDNAs. *Proc. Natl. Acad. Sci. U.S.A.* 96:9345-50.
- Neznanov, N., K. M. Chumakov, L. Neznanova, A. Almasan, A. K. Banerjee, and A. V. Gudkov. 2005. Proteolytic cleavage of the p65-RelA subunit of NF- κ B during poliovirus infection. *J. Biol. Chem.* 280:24153-24158.
- Palese, P., and M. L. Shaw. 2007. Orthomyxoviridae: the viruses and their replication, p. 1647-1689. In D. M. Knipe and P. M. Howley (ed.), *Fields virology*. Lippincott Williams & Wilkins, Philadelphia, Pa.
- Park, S., X. Yang, and J. G. Saven. 2004. Advances in computational protein design. *Curr Opin Struct Biol* 14:487-94.
- Paul, A. V., J. A. Mugavero, A. Molla, and E. Wimmer. 1998. Internal ribosomal entry site scanning of the poliovirus polyprotein: implications for proteolytic processing. *Virology* 250:241-253.
- Pelletier, J., and N. Sonenberg. 1988. Internal initiation of translation of eukaryotic mRNA directed by sequence derived from poliovirus RNA. *Nature.* 334:320-325.
- Pfister, T., and E. Wimmer. 1999. Characterization of the nucleoside triphosphatase activity of poliovirus protein 2C reveals a mechanism by which guanidine inhibits poliovirus replication. *J. Biol. Chem.* 274:6992-7001.
- Plotkin, J. B., H. Robins, and A. J. Levine. 2004. Tissue-specific codon usage and the expression of human genes. *Proc. Natl. Acad. Sci. U.S.A.* 101:12588-12591.
- Racaniello, V. R., and D. Baltimore. 1981. Cloned poliovirus complementary DNA is infectious in mammalian cells. *Science.* 214:916-9.
- Reed, L. J., and M. Muench. 1938. A simple method for estimating fifty percent endpoints. *Am. J. Hyg.* 27:493-497.
- Richardson, S. M., S. J. Wheelan, R. M. Yarrington, and J. D. Boeke. 2006. GeneDesign: rapid, automated design of multikilobase synthetic genes. *Genome Res.* 16:550-556.
- Robinson, M., R. Lilley, S. Little, J. S. Emteage, G. Yarrington, P. Stephens, A. Millican, M. Eaton, and G. Humphreys. 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nucl. Acids Res.* 12:6663-6671.
- Rothberg, E. 1985. wmatch: a C program to solve maximum-weight matching. [Online.]
- Rueckert, R. R. 1985. Picornaviruses and their replication, p. 705-738. In B. N. Fields, D. M. Knipe, R. M. Chanock, J. L. Melnick, B. Roizman, and R. E. Shope (ed.), *Fields virology*, vol. 1. Raven Press, New York, N.Y.
- Russell, C. J., and R. G. Webster. 2005. The genesis of a pandemic influenza virus. *Cell.* 123:368-371.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Sánchez, G., A. Bosch, and R. M. Pinto. 2003. Genome variability and capsid structural constraints of hepatitis A virus. *J. Virol.* 77:452-459.
- Savolainen, C., S. Blomqvist, and T. Novi. 2003. Human rhinoviruses. *Paediatr. Respir. Rev.* 4:91-98.
- Schwerdt, C., and J. Fogh. 1957. The ratio of physical particles per infectious unit observed for poliomyelitis viruses. *Virology* 4:41-52.

99

- Shimizu, H., B. Thorley, F. J. Paladin, K. A. Brussen, and V. Stambos et al. 2004. Circulation of type 1 vaccine-derived poliovirus in the Philippines in 2001. *J. Virol.* 78:13512-13521.
- Simonsen, L., T. A. Reichert, C. Viboud, W. C. Blackwelder, R. J. Taylor, and M. A. Miller. 2005. Impact of influenza vaccination on seasonal mortality in the US elderly population. *Arch. Intern. Med.* 165:265-272.
- Skiena, S. S. 2001. Designing better phages Bioinformatics. 17 Suppl 1:5253-61.
- Steinhauer, D. A., and J. J. Skehel. 2002. Genetics of influenza viruses. *Annu. Rev. Genet.* 36:305-332.
- Stephenson, I., and J. Democratis. 2006. Influenza: current threat from avian influenza. *Br. Med. Bull.* 75-76: 63-80.
- Svitkin, Y. V., G. A. Alpatova, G. A. Lipskaya, S. V. Maslova, V. I. Agol, O. Kew, K. Meerovitch, and N. Sonenberg. 1993. Towards development of an in vitro translation test for poliovirus neurovirulence. *Dev. Biol. Stand.* 78:27-32.
- Svitkin, Y. V., S. V. Maslova, and V. I. Agol. 1985. The genomes of attenuated and virulent poliovirus strains differ in their in vitro translation efficiencies. *Virology* 147:243-252.
- Talon, J., M. Salvatore, R. E. O'Neill, Y. Nakaya, H. Zheng, T. Muster, A. Garcia-Sastre, and P. Palese. 2000. Influenza A and B viruses expressing altered NS1 proteins: A vaccine approach. *Proc. Natl. Acad. Sci. U.S.A.* 97:4309-4314.
- Thompson, W. W., D. K. Shay, E. Weintraub, L. Brammer, N. Cox, L. J. Anderson, and K. Fukuda. 2003. Mortality associated with influenza and respiratory syncytial virus in the United States. *JAMA.* 289:179-186.
- Tian, J., H. Gong, N. Shang, X. Zhou, E. Gulari, X. Gao, and G. Church. 2004. Accurate multiplex gene synthesis from programmable DNA microchips. *Nature.* 432: 1050-1054.

100

- Tolskaya, E. A., L. I. Romanova, M. S. Kolesnikova, T. A. Ivannikova, E. A. Smirnova, N. T. Raikhlin, and V. I. Agol. 1995. Apoptosis-inducing and apoptosis-preventing functions of poliovirus. *J. Virol.* 69:1181-1189.
- Toyoda, H., J. Yin, S. Mueller, E. Wimmer, and J. Cello. 2007. Oncolytic treatment and cure of neuroblastoma by a novel attenuated poliovirus in a novel poliovirus-susceptible animal model. *Cancer Res.* 67:2857-64.
- van der Wert, S., J. Bradley, E. Wimmer, F. W. Studier, and J. J. Dunn. 1986. Synthesis of infectious poliovirus RNA by purified T7 RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* 78:2330-2334.
- Wahby, A. F. 2000. Combined cell culture enzyme-linked immunosorbent assay for quantification of poliovirus neutralization-relevant antibodies. *Clin. Diagn. Lab. Immunol.* 7:915-9.
- Wang, B., D. Papamichail, S. Mueller, and S. Skiena. 2006. Two Proteins for the Price of One: The Design of Maximally Compressed Coding Sequences Natural Computing. Eleventh International Meeting on DNA Based Computers (DNA11), 2005. Lecture Notes in Computer Science (LNCS), 3892:387-398.
- Zhao, W. D., and E. Wimmer. 2001. Genetic analysis of a poliovirus/hepatitis C virus chimera: new structure for domain II of the internal ribosomal entry site of hepatitis C virus. *J. Virol.* 75:3719-3730.
- Zhou, J., W. J. Liu, S. W. Peng, X. Y. Sun, and I. Frazer. 1999. Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. *J. Virol.* 73:4972-4982.
- Zolotukhin, S., M. Potter, W. W. Hauswirth, J. Guy, and N. Muzyczka. 1996. A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells. *J. Virol.* 70:4646-4654.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 42

<210> SEQ ID NO 1
<211> LENGTH: 2643
<212> TYPE: DNA
<213> ORGANISM: Human poliovirus 1

<400> SEQUENCE: 1

atgggtgctc aggtttcatc acagaaaatg ggcgcacatg aaaactcaaa tagagcgat	60
ggtggttcta ccattaatta caccaccatt aattattata gagattcagc tagtaacgcg	120
gcttcgaaac aggacttctc tcaagaccct tccaagtctca ccgagccat caaggatgtc	180
ctgataaaaa cagccccaat gctaaactcg ccaaacaatag aggcttgcgg gtatacgat	240
agagtactgc attaacact gggaaactcc actataacca cacaggaggc ggctaattca	300
gtagtcgctt atgggcgttg gcctgaatat ctgagggaca gcgaagccaa tccagtggac	360
cagccgacag aaccagacgt cgctgatgc aggtttata cgctagacac cgtgtcttgg	420
acgaaagagt cgcgagggtg gtgggtggaaat ttgcctgtat cactgaggaa catggactc	480
tttggccaaa atatgtacta ccactaccta ggttaggtccg ggtacaccgt gcatgtacag	540
tgtAACGCTT ccaaattcca ccaggggca cttagggtat tcggcgatcc agagatgtgt	600
ctggccgggg atagcaacac cactaccatg cacaccagct atcaaaatgc caatcctggc	660
gagaaaaggag gcactttcac gggtaacgttc actcctgaca acaaccagac atcacctgcc	720

-continued

cgcagggttct gccccgtgga ttacccctt ggaaatggca cgttgttggg gaatgcctt	780
gtgttcccgcc accagataat aaacatcacgg accaacaact gtgctacact ggtactccct	840
tacgtgaact ccctctcgat agatagtagt gtaaagcaca ataattggg aattgcaata	900
ttaccattgg ccccattaaa ttttgctagt gagtcctccc cagagattcc aatcacctt	960
accatagccc ctatgtgctg tgatgtcaat ggattaagaa acatcacccct gccacgctta	1020
cagggectgc cggcatgaa cacccttgtt agcaatcaat atcttactgc agacaacttc	1080
cagtcaccgt gtgcgctgcc tgaatttgcgt gtgacccac ctattgacat acccggtgaa	1140
gtaaagaaca tcatggatt ggcagaaatc gacaccatgaa ttcccttgc cttaaatgtgc	1200
acaaaaaaaga acaccatgaa aatgtatagg gttcggttac gtaacaaacc acatacagac	1260
gatcccatac tctgcctgtc actctctcca gcttcagatc ctatgggtc acatactatg	1320
cttggagaaa tcctaaatca ctacacacac tggcaggat ccctgaagtt cacgttctg	1380
ttctgtggat tcatgtggc aactggcaaa ctgttgggtc catacgcgcc tcctggagcc	1440
gaccaccaa agaagcgtaa ggaggcgatg ttgggaacac atgtgtatcg ggacatagga	1500
ctgcagtcct catgtactat ggtatgcctt tggattgca acaccacgtt tcggcaaaacc	1560
atagatgata gtttaccgaa aggccggatac atcagcgtct tctaccaaac tagaatagtc	1620
gtccctctt ccacacccag agagatggac atccttgggtt ttgtgtcagc gtgtaatgac	1680
ttcagcgtgc gcttgggtcg agataccaca catatagac aaaaagcgct agcacagggg	1740
ttaggtcaga tgcttggaaatg catattgac aacacagtcc gtgaaacggg gggggcggca	1800
acatcttagag acgctctccc aaacactgaa gccagtgac caacacactc caaggaaatt	1860
ccggcactca ccgcgtggaa aactggggcc acaaattccac tagtcccttc tgatacagt	1920
caaaccacag atgttgtaca acataggtaa aggtcagatg ctagcataga gtcttttttc	1980
gcgcgggggtg catgcgtgac cattatgacc gtggataacc cagcttccac cacgataaag	2040
gataagctat ttgcgtgtg gaagatcaat tataaagata ctgtccaggat acggaggaaa	2100
ttggagttct tcaccttattc tagatttgcgt atggaaactt cctttgtggt tactgcaaat	2160
ttcaactgaga ctaacaatgg gcatgcctt aatcaagtgtt accaaatttat gtacgtacca	2220
ccaggcgctc cagtgcggca gaaatggac gactacacat ggcaaaacccat atcaaattcca	2280
tcaatctttt acacctacgg aacagctcca gcccggatct cggtaccgtt tggtggatt	2340
tcgaacgcct attcacactt ttacgacggt tttccaaatg taccactgaa ggaccagtcg	2400
gcagcactag gtgactccct ttatgggtca gcatcttcaa atgacttcgg tattttggct	2460
gttagagtag tcaatgtatca caacccgacc aaggtcacctt ccaaaatcag agtgtatcta	2520
aaacccaaac acatcagatgt ctgggtcccg cgtccaccgaa gggcagtggc gtactacggc	2580
cctggagtggtt attacaagga tggtaacgctt acaccctctt ccaccaagga tctgaccaca	2640
tat	2643

<210> SEQ ID NO 2
 <211> LENGTH: 7521
 <212> TYPE: DNA
 <213> ORGANISM: Unknown
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 2

ttaaaaacagc tctgggggttacccaccc agaggccac gtggcggcta gtactccgg	60
attgcggtaaccccttata ctcccttccc gtaacttaga cgcacaaaac	120

-continued

caagttcaat agaagggggt acaaaccagt accaccacga acaagcaccc	180
ggtgatgtcg tataactgc ttgcgtgggt gaaagcgacg gatccgttat ccgcttatgt	240
acttcgagaa gcccagtagcc acctcgaaat ctgcgtcg ttcgcgtcag cactcaaccc	300
cagagtgttag cttaggctga tgagtctgga catccctcac cggtgacggt ggtccaggt	360
gcgttggcgcc cctacctatg gctaacgcca tgggacgcta gttgtgaaca aggtgtgaag	420
agcctattga gctacataag aatctccgg cccctgaatg cggctaatcc caaccccgaa	480
cgaggtggtc acaaaccagt gattggcctg tcgtaacgca caagtcgtg gcggAACCGA	540
ctactttggg tgcgtgtt tcctttatt ttattgtggc tgcttatgtt gacaatcaca	600
gattgttatac ataaagcgaa ttggattggc catccgtga aagttagact cattatctat	660
ctgtttgtc gatccgtcc attgagtgtg ttactctaa gtacaatttc aacagttatt	720
tcaatcagac aattgtatca taatgggtgc tcagggttcg tcgcaaaaag taggtgcgc	780
tgaaaactcg aatcggtcgat atgggtcgac gacgataat tatacgacga taaattatta	840
tcgtgattcg gctgtataatg cggcttcgaa acaagatttt tcgcaagatc cgctgaaatt	900
tacggaaaccg ataaaatgt tattaataaa aacggcgccg atgttaattt cgccgaaat	960
agaagcgtgt gtttattcgg atcggttattt acaattaaacg ttaggttaattt cgacgataac	1020
gacgcaagaa cggcgttaattt cggtagtgcgat gtatggcgt tggccggagt atttacgtga	1080
ttcggaaagcg aatccggtag atcaaccgac ggaaccggac gtcgccccgt gtcgttttta	1140
tacgttagat acggatcgat ggacgaaaga atcggtcgat tgggtggta aattaccggaa	1200
tgcgttacgt gatatgggtt tattttgtca aaatatgtat tatcattatt taggtcggtt	1260
gggttatacg gtacatgtac aatgtatgc gtcgaaattt catcaagggtt cgtaggtgt	1320
atttcggttcc cggaaatgt gtttagcggtt tgactcgaat acgacgacga tgcatacg	1380
gtatcaaaat gcgaatccgg gtggaaaagg tggtagtgc ttacgcccga	1440
taataatcaa acgtcgccgg cgcgtcggtt ttgtccgtt gattattttat taggtatgg	1500
tacgttattt gggaaatgcct ttgttattcc gcatcaataa ataaatttgc gtacgaaataa	1560
tttgtgegacg ttagtattac cgtatgtaaa ttgcgttatcg atagattcga tggtaaaaca	1620
taataattgg ggtatagcaa tattaccgtt agcgccgtt aattttcggtt cggaatcg	1680
gccggaaataa cggataacgt taacgtatgc gccgatgtgt tgtgaatttta atgggttgc	1740
taatataacg ttaccgtt tacaagggtt accggtaatg aatacgccgg gtcgaaatca	1800
gtatccaacg cggataattt ttcaatcgcc gtgtgcgtt ccggaaatttg atgtacg	1860
gccgatagat ataccgggtt aagtaaaaaa tatgtatggaa ttagcgaaa tagatacgat	1920
gataccgtt gatttacgtt cggacaaaaaa aaatacgatg gaaatgtatc gtgtacgtt	1980
atcgatggaa cggcataacgg atgatccgtt attatgttta tcgttatcgcc cgccgtcg	2040
tccttaggttta tcgcatacga tgtaggtga gatattaaat tattatacgc attggccgg	2100
ttcgtaaaaa ttacgtttt tattttgtgg ttgcgtatgcg ggcacgggtt aatttattgt	2160
atcgatgtcg cggccgggtt cggatccggc gaaaaaacgt aaagaagcga tgtaggtac	2220
gcgtgtataa tgggtatatacg tttacaatc gtcgtgtacg atggtagtac cgtaggtatc	2280
gaatacgacg tatcgtaaa cgtatgtacg ttcgtttacg gaaggtgggtt atatatcggt	2340
attttatcaa acgcgtatag tagtaccgtt atcgacgccc cgtagaaatgg atatattagg	2400
ttttgtatcg gcgtgtatgcg attttcggt acgttttattt cgtgatacga cgcatataga	2460

-continued

acaaaaagcg ctagcacaag gtttaggtca aatgttagaa tcgatgatag ataatacggt	2520
acgtgaaacg gtaggtgcgg ctacgtcgcg ttagtgcgtta ccgaataacgg aagcgtcggg	2580
tccgaegcac tcgaaagaaa tacccggcggtt aacggcggtt gaaacgggtt cgacgaatcc	2640
actagtccccg tcggatacgg tacaaacggc tcatgttagt caacatcggtt cgcggtcgga	2700
atcgtcgata gaatcggttt ttgcgcgtgg tgctgtgtt acgataatgtt cggttagataa	2760
tccggcgctc acgacagaata aagataaattt atttgcgggtt tgaaaataaa cgtataaaga	2820
tacggtacaa ttacgtcgta aattagaattt tttcacgtat tgcgtttt atatggattt	2880
aacggttgcgta gtaacggcgaa atttacggaa aacgaataat ggtcatcggt taaatcaagt	2940
atatcaaata atgtacgtac cgccgggtgc gcccgtaccc gaaaaatggg atgattatac	3000
gtggcaaacg tcgtcgatatt ttacacgtat ggtacggcgcg cggcgctat	3060
atcggtaccg tatgttaggtt tatctaattgc gtatttcgtat ttttatgtt gtttctcgaa	3120
agtaccgtta aaagatcaat cggcgccgtt aggtgattcg ttatatgggtt cggcgctgtt	3180
aatgattttt ggtatatttgcggtt agttaatgtt cataatccga cgaagggtcac	3240
ctcgaaaata cgtgtatatt taaaaccgaa acatatacgt gtatgggtgc cgcgtccgccc	3300
gcgtgcggta gcgtattatgtt gtcgggtgtt agattataaa gatggtacgt taacggcggtt	3360
atcgacgaaa gatttaacga cgtagtgattt cggacaccaa aacaaacggc tgcgtactgc	3420
aggttacaaa atttgcactt accacttggc cacttcaggat gatttgcaaa acgcgtgtt	3480
cgtcatgtgg agtagagacc tcttagtgcg agaatcaaga gcccaggcgc cccgttcaat	3540
cgcaagggtgc aattgcacgc caggggtgtt ctactgcgag tctagaagga aataactaccc	3600
agtatccctc gttggcccaa cgttccagta catggaggctt aataactattt acccagctgtt	3660
gttaccaggcc catatgtca ttggccatgg atttcgtatctt ccaggggattt gtgggtggcat	3720
actcagatgtt caccacgggg tgatagggat cattactgtt ggtggagaag gggtgggtgc	3780
attttcagac atttagagact tgcgtgccta cgaagaagaa gccatggaa aaggcctcac	3840
caattacata gagtcacttgc gggccgcattt tggaaatggaa ttactcgtt agattagcgtt	3900
caaaataaca gagttgacca atatgggtac cagtaccatc actgaaaagc tacttaagaa	3960
cttgcgtcaag atcatatcctt cactgttat tataacttagt aactatgtt aacaccacac	4020
agtgcgtgcgtt accctggccc ttcttgggtt tgatgcgttca ccatggcgtt ggcttagaaa	4080
gaaagcatgc gatgttctgg agataccctt tgcgtcaagt caaggtgaca gttgggttgc	4140
gaagtttactt gaagcatgc acgcgttaccc gggactggag tgggtgttcaaa aaaaaatctc	4200
aaaaatttattt gattggctca aggagaaaat tatcccacaa gcttagagata agttggattt	4260
tgtacaaaaa ctttagacaac tagaaatgtt gggaaaccaa atctcaacta tacaccaatc	4320
atgccttgcgtt caggaacacc accggaaatttctt attcaataat gtcgtatgtt tatccatcca	4380
gttcaagagg tttgccttc ttacgtcgat ggaagccaa agaatacaga aactcgagca	4440
tactattaac aactacatac agttcaagag cttttttttt attgtttttttt tatgtttgtt	4500
agtacatggc agccccggaa caggtaaatc tgcgtgcgtt aacctgtatgtt ctacgtccat	4560
agctgaaaga gaaaacacgtt ccacgtactc gttttttttt gatccatcac acttcgacgg	4620
atacaaacaa cagggtgtgg tgattatggaa cgacgttcaat cttttttttt atgggtgcgg	4680
catgaagctg ttctgtcaga tggatcaac agtgggtttt ataccacccaa tggcatccctt	4740
ggaggagaaa ggaatccctgtt ttacttcaaa ttacgttctt gcatccacaa actcaagcgtt	4800
aatttccccccccc cccactgtgg cacacgtgtt cgcgttagcc accggcttttgcgat	4860

-continued

ggacattcag gtcatgaatg agtattctag agatggaaa ttgaacatgg ccatggctac 4920
 tgaatgtgt aagaactgtc accaaccgc aaactttaag agatgtgtc cttagtgt 4980
 tggtaaggca attcaattaa tggacaaatc ttccagagt agatacagta ttgaccagat 5040
 cactacaatg attatcaatg agagaaacag aagatccaac attggcaatt gtatggagc 5100
 ttgtttcaa ggaccactcc agtataaaga cttgaaaatt gacatcaaga cgagtcccc 5160
 tcctgaatgt atcaatgact tgctccaagc agttgactcc caggaggtga gagattactg 5220
 tgagaagaag ggttgatag ttaacatcac cagccaggtt caaacagaaa ggaacatcaa 5280
 cagggcaatg acaattctac aacgcgtgc aacccgcgc gcagtggctg gagttgtcta 5340
 tgtcatgtat aaactgtttg ctggacacca gggagcatac actggttac caaacaaaaa 5400
 acccaacgtg occaccattc ggacagcaaa ggtacaagg ccagggttcg attacgcgt 5460
 ggctatggct aaaagaaaaca ttgttacagc aactactagc aaggagagt tcactatgtt 5520
 aggagtcac gacaacgtgg ctatccc aacccacgcg tcacctggtg aaagcatgt 5580
 gatcgatggc aaagaagtgg agatcttgg tgccaaagcg ctcgaagatc aagcagggac 5640
 caatcttcaa atcactataa tcactctaa gagaatgaa aagttcagag acattagacc 5700
 acatataacct actcaaatac ctgagacaaa tggatggatc ttgatgtga acactagcaa 5760
 gtaccccaat atgtatgttc ctgtcggtgc tgtgactgaa cagggatatc taaatctcg 5820
 tggcgcccaa actgtcgta ctctaatgta caacttcca accagagcag gacagtgtgg 5880
 tggagtcatc acatgtactg gggaaatgtcat cgggatgtatc gttggggaa acgggttcaca 5940
 cgggttgca gcggccctga agcgatcata ctgcactcg agtcaaggaa aaatccagtg 6000
 gatgagacct tcgaaggaag tgggatatcc aatcataat gccccgtcca aaaccaagct 6060
 tgaacccagt gcttccact atgtgtttga aggggtgaag gaaccagcag tcctcaactaa 6120
 aaacgatccc aggcttaaga cagactttga ggaggcaatt ttctccaatg acgtggtaa 6180
 caaaattact gaagtggatg agtacatgaa agaggcgtt gaccactatg ctggccagct 6240
 catgtcaacta gacatcaaca tagacaatgt tgcttggatgatgtt atggactgt 6300
 tggtctagaa gcacttgatt tgccaccag tgctggctac ctttatgttag caatggaaa 6360
 gaagaagaga gacatcttga acaaacaac cagagacact aaggaaatgc aaaaactgct 6420
 cgacacatat ggaatcaacc tcccactgtt gacttatgtt aaggatgaaatc ttagatccaa 6480
 aacaaagggtt gggcgggaa aatccagatt aattgaagct tctatgttga atgactcagt 6540
 ggcaatgaga atggcttttgg ggaacctata tgctgtttt cacaaaaacc caggagtgt 6600
 aacaggttca gcgtgggtt gcgatccaga tttgtttgg agcaaaatcc cggtattgt 6660
 ggaagagaag ctgtttgtt ttgactacac agggatgtatgat gcatctctca gcccgtctg 6720
 gttcgaggca ctaaagatgg tgcttgagaa aatcggttcc ggagacagag ttgactacat 6780
 cgactaccta aaccactcac accacctgtt caagaataaa acatactgtt tcaagggcgg 6840
 tatgccatct ggctgtctcg gcacttcaat tttaactca atgattaaca acttgattat 6900
 caggacactc ttactgaaaa cctacaaggcatgat tttttttt gaccacatca aatgtatgc 6960
 ctatggatgatgtt gatgttgcctt cttcttaccc ccatgttgcgtt gacgttgcgtt ccctagccca 7020
 atcaggaaaa gactatggac taactatgtac tccagctgac aaatcgctt catttggaaac 7080
 agtcacatgg gagaatgttca cattttttttt gatgttgcgtt agggcagacg agaaataaccc 7140
 atttcttattt catccagttaa tgccatgaa ggaaattcat gaatcaatggatggacaaa 7200

-continued

agatcctagg aacactcagg atcacgttcg ctctctgtgc ctttagctt ggcacaatgg	7260
cgaagaagaa tataacaaat tccttagctaa aatcaggagt gtgccaattg gaagagctt	7320
attgctccca gagtactcaa cattgtaccg cggttgctt gactcattt agtaacccta	7380
cctcagtcga attggattgg gtcatactgg tgtagggta aattttctt taattcgag	7440
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa	7500
aaaaaaaaaa aaaaaaaaaa a	7521

<210> SEQ ID NO 3

<211> LENGTH: 7521

<212> TYPE: DNA

<213> ORGANISM: Unknown

<220> FEATURE:

<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 3

ttaaacacgc tctggggtt tacccacccc agaggcccac gtggcggcta gtactccggt	60
attgcggta ccttgtacgc ctgtttata ctcccttccc gtaacttaga cgcacaaaac	120
caagttcaat agaagggggt acaaaccagt accaccacga acaagcaccc ctgtttcccc	180
ggtgatgtcg tatagactgc ttgcgtggtt gaaagcgcac gatccgttat ccgcttatgt	240
acttcgagaa gcccagtagc acctcggaaat ctgcgtcg ttgcgtccag cactcaaccc	300
cagagtgttag cttaggctga tgagtctgga catccctcac cggtgacggt ggtccaggt	360
gcgttggcgcc cctacctatg gctaacgcca tgggacgcta gttgtgaaca aggtgtgaag	420
agcctattga gctacataag aatcctccgg cccctgaatgc cggtcaatcc caacctcgga	480
gcaggggtgc acaaaccagt gattggcctg tcgtaacgcg caagtccgtg gcggAACCGA	540
ctactttggg tgcgtgtt tcctttatt ttattgtggc tgcttatgtt gacaatcaca	600
gattgttatac ataaagcgaa ttggattggc catccggtaa aagttagact cattatctat	660
ctgtttgtc gatccgtcc attgatgtgt ttactctaa gtacaatttc aacagttatt	720
tcaatcagac aattgtatca taatgggtgc tcagggttcg tctcagaagg taggtgtca	780
cggaaatagc aacaggcat acggatcatc cacaatcaac tatacaacaa tcaactacta	840
ccgggactcc gcctccaaatg cagttcgaa acaagattt agccaggatc catctaaatt	900
tactgaacca attaaagacg tgctcatca gactgcaccc atgttgaata gccctaacat	960
tgaaggctgt ggctactcg accgggtgtc tcagctgacc ctggtaatt caaccatcac	1020
aacgcacaa gctgcaaaaca gctgtggc atacggaaagg tggccagagt acttaagaga	1080
ttctgaggca aaccctgtcg atcaaccaac cgagccggac gtgcggcgt gtcgattcta	1140
caccctcgat actgtaagct ggaccaagga atcgcgaggat tgggtggtaa aacttccaga	1200
cgccctacgc gatatgggtt tgggtggtaa gaacatgtat tattttttt tgggacgctc	1260
tggatataca gtccacgtgc aatgcaatgc ttcaaaatgtt catcaggaggtt ctgggggtgt	1320
tttcgcgggtc cggaaatgt gcttagcagg agactcaaat acgaccacaa tgcatacttc	1380
ttaccagaac gcaaaaccccg gagaaaaggcg cggtaatgtt accgggaccc ttacccaga	1440
taataatcaa acctcccccag caagacgttt ttgtccaggat gactattgt tgggcaacgg	1500
tacactacta gggaaatgcct tcgtatcc tcgtatccatc atcaatcttta gaactaataa	1560
ttgcgcacta ctgcgtgtc catatgtaaa ttctctgtcc atcgactcga tgggtggtaa	1620
taacaactgg gggatagcaa tattacctctcg cggccctctg aacttcgcgt cagaatcaag	1680
tccggaaatc ccgataacac ttactatcgc tcccatgtgt tgcgagttca acgggctcag	1740

-continued

gaatataaca ctacccaggc tgcaagggt accagttatg aatacaccag gatctaacc	1800
gtacttaacc gccgataact tccaatcccc atgcgcctta ccagagttcg atgtcacgcc	1860
gccccatcgat atccctgggg aggtgaaaa tatgatggag cttgcccaga ttgatacaat	1920
gataccattc gatctgtccg caactaaagaa aaataacaatg gagatgtacc gtgtgagact	1980
ttccgataag cctcacaccg acgacccaat tctgtgtctc agtctatctc ctgcaagtga	2040
cccttaggctt agccacacca tgctggcgta gatattgaac tattataccc actgggcggg	2100
cagtcttaaa ttcacccctt tattttgtgg gtcaatgtatg gctacgggta agttgttagt	2160
aagctacgctt ccaccaggcg cagatccctcc gaagaaaagg aaagaagctt tgcttgggac	2220
ccacgtcatt tgggatattt gtttacaatc ttctgtgcacc atgggtggtc cttggatatc	2280
aaatacggacc tacagacaga caatcgacga ctcccttacg gagggtgggtt atatatcagt	2340
gttttatcag acccgatattt tggtgccact gtctacccca cgggaaatgg atattctggg	2400
attcgtctcc gcctgcaacg acttctcagt caggctctta agggacacga cccacattga	2460
acagaaggcg otacgacaag gactgggtca aatgttgagatg tcaatgtatg ataataccgt	2520
aagggagacc gtaggagctg ccacccatcg ggatgcactg ccgaataacag aggcacatcagg	2580
gccccacccat tcaaaagaga tcccacgtct gacagctgtt gagaacctgtt caaccaaccc	2640
actagtccca tcggacactg ttccagacacg gcacgtgggtt cagcacagaa gtatccgaa	2700
atccctegatt gaaagctttt tcggcagagg cgccctgttata acaataatga cggtagacaa	2760
tcccgcagaat actaccaaca aagacaaact gttcgctgtt tggaaaataaa catacaagga	2820
caccgtacaa ctgagaagaa agcttgaatt ttttacttac agcagggttcg acatggagtt	2880
aacattcgtt gtgaccgctt acattcaccga aaccaataac gggcacgccc tgaaccaggt	2940
ttatcagata atgtacgtac cgccgggttc ccccgatccc gaaaagtggg atgattatac	3000
ttggcagact agctccaaacc ctccatatt ctatacttat ggcacccgcgc cggcgagaat	3060
ttcagttcca tacgtggaa tatcaaatgc atactctcat ttctatgtt gcttcctcaa	3120
gggtgcctta aagaccaat cageggccctt gggagattcg ttgtacgggg ccgcctcctt	3180
gaacgatttt gggatcctag cagtgggggtt ggtgaacgac cataatccaa cgaaggtcac	3240
cagtaagata cgggtctact tgaaggctaa gcatattcg ttttttttttcaagaccc	3300
acgcgcgcgc gcatattatg gacccgggtt cgactataaa gacgggaccc ttggccact	3360
gagttacaaa gacccatcaa cttacggatt cggacaccaa aacaaaggcg ttgtacactgc	3420
aggttacaaa atttgcactt accacttggc cactcaggat gatggcaaa acgcagtgaa	3480
cgtcatgtgg agtagagacc ttcttagtccatc agaatcaaga gcccaggggca ccgattcaat	3540
cgcagggttc aattgcacccg cagggggttca ctactgcgag tctagaaggaa aataactaccc	3600
agtatcccttc gttggcccaa cgttccatca catgggggtt aataactatt acccagcttag	3660
gtaccaggccatca ttggccatgg attcgcattt ccagggttccatggtggcat	3720
actcagatgtt caccacgggg tggatggat cattactgtt ggtggagaag ggttggttgc	3780
attttcagac atttagactt ttttttttttca gcaatggaaac aaggccctcac	3840
caattacata gaggacttgg gggccgcatt tggaaatggaa ttttttttttca gatggcgaa	3900
caaaataaca gagttgacca atatgggtac cagttaccatc actggaaaac tacttaagaa	3960
cttgcgttcaatcactatccctt cacttagtttataacttagg aactatgttcaac accaccacaa	4020
agtgcgttcaaccctggccc ttcttgggtt gtttttttttca ccatggcagt ggcttagaaa	4080

-continued

gaaagcatgc gatgttctgg agatacccta tgcaccaag caaggtgaca gttgggtgaa	4140
gaagtttact gaagcatgca acgcagctaa gggactggag tgggtgtcaa acaaaaatctc	4200
aaaattcatt gattggctca aggagaaaat tatcccacaa gctagagata agttggatt	4260
tgtacaacaa cttagacaac tagaaatgct ggaaaaccaa atctcaacta tacaccaatc	4320
atgccctagt caggaacacc agggaaattctt attcaataat gtcagatggt tatccatcca	4380
gtctaagagg tttgccctc tttacgcagt ggaagccaa agaatacaga aactcgagca	4440
tactattaac aactacatac agttcaagag caaacaccgt attgaaccag tatgtttgct	4500
agtacatggc agccccggaa caggtaaatc tgttagcaacc aacctgatgt cttagccat	4560
agctgaaaga gaaaacacgt ccacgtactc gctacccccc gatccatcac acttcgacgg	4620
atacaaacaac caggggagtgg tgattatgga cgacctgaat caaaacccag atggcgga	4680
catgaagctg ttctgtcaga tggtatcaac agtggagttt ataccaccca tggcatccct	4740
ggaggagaaa ggaatcctgt ttacttcaaa ttacggtcta gcatccacaa actcaagcag	4800
aatttcccccc cccactgtgg cacacagtga cgcgttagcc aggcgccttg cgttcgacat	4860
ggacattcg gtcatgaatg agtattctag agatggaaat ttgaacatgg ccatggctac	4920
tgaaatgtgt aagaactgtc accaaccacgaa aaactttaag agatgtgtc cttagtgt	4980
tggtaaggca attcaattaa tggacaaatc ttccagagtt agatacagta ttgaccagat	5040
cactacaatg attatcaatg agagaaacag aagatccaaatc attggcaattt gtagggagc	5100
tttgttcaaa ggaccactcc agtataaaga cttgaaaattt gacatcaaga cgagtcccc	5160
tcctgtatgt atcaatgact tgcgttcaacg agttgactcc caggaggtga gagattactg	5220
tgagaagaag ggttggatag ttaacatcac cagccaggtt caaacagaaa ggaacatcaa	5280
cagggcaatg acaattctac aagcggtgac aaccttcgccc gcagtggtcg gagttgtct	5340
tgtcatgtat aaactgtttt ctggacacca gggagcatac actggtttac caaaacaaaa	5400
accccaacgtg cccaccatc ggacagcaaa ggtacaaggaa ccagggttgc attacgcagt	5460
ggctatggct aaaagaaaaca ttgttacagc aactactagc aaggagagt tcactatgtt	5520
aggagtccac gacaacgtgg ctattttacc aacccacgtc tcacctgggtg aaagcatgt	5580
gatcgatggc aaagaagtgg agatcttgg tggccaaagcg ctcgaagatc aagcaggAAC	5640
caatcttggaa atcactataa tcactctaaa gagaatgaa aagttcagag acattagacc	5700
acatataacct actcaaatac ctgagacaaa tggatggatc ttgatgtca acactagcaa	5760
gtaccccaat atgtatgttc ctgtcggtgc tggactgaa cagggtatc taaatctcg	5820
tgggcgc当地 actgtcgta ctctaatgtca caactttcca accagagcag gacagtgtgg	5880
tggagtcatac acatgtactg ggaaaagtcat cgggatgtatc gttgggtggaa acgggtcaca	5940
cgggtttgca gggccctga agcgatcata ctgcactcag agtcaagggtg aaatccagt	6000
gatgagaccc tgcgaaaggaaat tgggtatcc aatcataat gccccgttcaaa aaaccaagct	6060
tgaacccagt gctttccact atgtgttgc aggggtgaaag gaaccagcag tcctcaactaa	6120
aaacgatccc aggcttaaga cagactttga ggaggcaattt ttctccaatg acgtgggtaa	6180
caaaaattactt gaagtggatg agtacatgaa agaggcagta gaccactatg ctggccagct	6240
catgtcacta gacatcaaca tagaacaaat gtgtttggag gatgtccatgt atggcactga	6300
tggcttagaa gacttgcatt tgcgttccatc tgcgtggatc ctttatgttag caatggaaa	6360
gaagaagaga gacatcttgc acaaacaac cagagacact aaggaaatgc aaaaactgct	6420
cgacacatata ggaatcaacc tccactggt gacttgcata aaggatgaaat ttagtccaa	6480

-continued

aacaaagggt gaggcagggga aatccagatt aattgaagct tcttagtttga atgactcagt	6540
ggcaatgaga atggcttttg ggaacctata tgctgccttt cacaaaaacc caggagtgtat	6600
aacagggttca gcagtggggt gcgatccaga ttgttttgg agcaaaaattc cggttattgtat	6660
ggaagagaag ctgtttgcct ttgactacac agggtatgtat gcatctctca gccctgcttg	6720
gttcgaggca ctaaagatgg tgcttgagaa aatcggttcc ggagacagag ttgactacat	6780
cgtactaccta aaccactcac accacactgtt caagaataaa acatactgtt tcaaggggcg	6840
tatgccatct ggctgcttag gcacttcaat tttaactca atgattaaca acttgattat	6900
caggacactt ttactgaaaa cctacaaggg catagatttta gaccacctaa aaatgattgc	6960
ctatggtgat gatgtatgtt cttccctaccc ccatgaagtt gagcgttagtc tcctagcccc	7020
atcaggaaaa gactatggac taactatgtac tccagctgac aaatcagctt catttggaaac	7080
agtcacatgg gagaatgtaa cattttgtttaa gagattttc agggcagacg agaaataccc	7140
atttcttattt catccagtaa tgccaatgaa ggaaattcatt gaatcaattt gatggacaaa	7200
agatccctagg aacactcagg atcacgttgc ctctctgtgc cttagttt ggcacaatgg	7260
cgaagaagaa tataacaat tccctagctaa aatcaggaggat gtgccttgcctt gaagacgttt	7320
attgctccca gagtactcaa cattgttcccg ccgttggctt gactcatttt agtaaccctt	7380
cctcagtcga attggattgg gtcatactgg tggatgggtt aattttctt taattcgag	7440
aaaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa	7500
aaaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa aaaaaaaaaaa	7521

```
<210> SEQ ID NO 4
<211> LENGTH: 2643
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: synthetic
```

<400> SEQUENCE: 4
atgggagctc aggtgtcatc ccaaaaagta ggccgcacacg aaaactcgaa tcgggcatac 60
ggatcttagta cgattaacta tactacaatc aattactata gggactccgc ctctaaccgc 120
gcttcgaaac aggactttc gcaagacccct agcaaggta ccgaaccgat taaggacgtg 180
tttgcataaa ccgcacccat gcttaactca cctaacatag aggcacgtcg 240
agggtgttgc aactgacact cggcaatagt acgattacga cacaggaagc cgctaatagc 300
gtatcgcat acggaagggtg gcgcgaatac ctttagagact ccgaagcgaa tccagtcgt 360
cagcctaccg aacctgacgt cggccatgt cggtttaca cactcgatac cgtgtcatgg 420
actaaggaaat cgcgagggtg gtggtgaaaa ctgccagacg cattgcgcga tatgggttg 480
ttcggacaga atatgtacta tcactatctc ggaagatccg ggtataccgt acacgtgcaa 540
tgttaacgcct ctaagttca ccagggagcg ttaggcgtat tcgcagtgcc agagatgtgt 600
ctagccggag actccaatac gactactatg catacctcat accaaaacgc taaccaggc 660
gaaaaggggg ggacattac cggAACATTC acacccgata acaatcagac atccccgct 720
agacgggtt gcccagtcga ctatctactc ggaaacggta cactgttagg gaatgccttc 780
gtatcccac accagataat taacttacgg actaacaatt gcgcaaccct agtggccaa 840
tacgttaact cactgtcaat cgatagtgatg gtgaaacata acaattgggg gatcgcaata 900
ttaccgttag cgccactgaa ttgcgcacg gaatcgatc ctgagatacc gattaccctt 960

-continued

acaatgcac ctatgtgttg cgagttcaac ggattgcgta atataaccct accacgggttgc	1020
cagggggtgc ccgttatgaa tacccccaggc tctaaccat accttaccgc cgataattc	1080
caatccctt gcgcaactgcc agagttcgac gtaaccctc caatcgacat acccggcgag	1140
gttaagaata tgatggagtt agccgaaatc gatactatga taccgttcga tctatccgct	1200
acgaaaaaga atactatgga gatgtatcgat gtgagattgt ccgataagcc acataccgac	1260
gatccgatac tgtgtctgtc actgtcaccgc ecagcgatc ctaggttgc ccataacaatg	1320
ttaggggaga tactgaatta ctataccat tggcccggtt gtctgaaatt cacattctg	1380
ttttggggat ctatgtatggc gacccggaaag ctgttagtgt catacgctcc accccggagcc	1440
gatccaccta aaaaacgcaa ggaagcgatc ctccgtacac acgtgatatg ggatatcgga	1500
ctgcaatcgat catgtactat ggtgtgtcca tggatatcgat atacgactta tagacagac	1560
atcgacgata gctttaccga gggggggat attagcgtat tctatcagac acgtatcgta	1620
gtgccactgt caacccttag agagatggac atactcgat tcgtatccgc atgtaacgac	1680
tttagcgtga gactgttacg cgatactacc catatcgaa acggaaaggcgt agcacagggg	1740
tttagggcaaa tgcttagatgc aatgtatcgat aatactgtac gcgaaaccgt aggccgcgc	1800
accagtaggg acgcactacc gaataccgaa gcgagcggac ctaccactc taaagagata	1860
ccggcactaa ccggccgtcgaa aaccggagcg actaaccacc tagtccccatc cgataaccgt	1920
caaaccagac acgttagtgca acatcggtct agatccgagt catcaatcgat atccctttc	1980
gctagaggcg catgcgttac gattatgacc gtgcataacc cagcgtcaac gactaacaaa	2040
gacaaattgt tcgcccgtatg gaaaattacc tataaggata ccgtgcattt ggcacgtaaa	2100
ctagagttct ttacataactc tagattcgat atggagctt cattcgtagt gaccgctaac	2160
tttaccgaga ctaataacgg tcacggccctt aatcagggtt atcagattat gtacgtaccg	2220
ccaggcgcac cagtggccgaa aaaatgggac gactatacat ggcagactag ctctaatccg	2280
tcaattttctt atacataacgg taccgcacca gctaggatta gcgtgcata cgtcggtata	2340
tgcgaaacgctt actcacactt ttacgacgga ttctctaaag tgccacttaa ggaccaatcc	2400
gcccgcactag gcgatagcctt atacggagcc gctagtttca acgatttccg tataactcgcc	2460
gttagggtcg tgaacgacca taaccctact aaggtcacctt ctaagattag ggtgtatctt	2520
aaggcctaaggc atattagggt gtgggtgtctt agaccgccta gagccgttgc gtattacgg	2580
ccaggcgtcg actataagga cggtacattt accccactgt caacgaaaga ccttacgacg	2640
tac	2643

<210> SEQ ID NO 5
 <211> LENGTH: 2643
 <212> TYPE: DNA
 <213> ORGANISM: Unknown
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 5	
atggggagctc aagtatcttc acaaaaagtt ggtgctcatg aaaattcaaa cagagcatat	60
ggttcttcaa caataaaacta caccacaata aattattaca gggacagtgc ttcaaatgc	120
gcttcgaaac aagatttttc tcaagatctt tccaagttca cagagccat caaagatgtt	180
ttaataaaaaa cagcacccat gctgaacgc cccaaacattt aagcatgtgg ttatagtgac	240
agagttctac agctcacgtt gggaaattca acaataacaa ctcaagaagc agcaaatct	300
gtgggtggctt atggtcgttg gccagagtat ctgcgggact cggaggcaaa tcctgtggac	360

-continued

cagccaacag agccggacgt cgctgcctgc cgcttctaca cgtagacac ggtctcctgg	420
accaaggagt cgcgaggctg gtggtgaaa cttccagatg ctctgcccga catggcctc	480
tttggcaga acatgtacta tcattattt ggaagaagtg getacacggt ccatgtcag	540
tgtaatgcataa ccaagttca tcaaggtgcc ctggcggtgt tgccggttcc tgagatgtgt	600
ttggctgggg acagcaacac caccaccatg cacacgtctt accaaaaatgc aaatccagga	660
aaaaaaagggtg gcacccatcac gggcaccttc actccagaca acaaccagac ctcgcggcg	720
cgccgcttgc gcccagtaga ttatctttttt gggaaatggca cgctgtggg gaatgccttt	780
gtatccctc atcaaataat aaatttgagg accaacaact gtgccacctt ggttttccct	840
tatgtcaaca gcctctccat tgactccatg gtgaagcaca acaactgggg aatagcaata	900
ttaccccttgg cacctttaaa ttttgcctcg gagagttctc cagagatccc catccaccc	960
accatcgccgc ogatgtgctg cgagttcaat gggctgagga acatccaccc gcccaggcta	1020
caagggtttc ctgtcatgaa cacaccaggta tcaaataatcgtt atttaaacgc agacaacttc	1080
cagtcctccat gtgtttggcc agaatttgat gtcaactccac caatagatat tccaggagaa	1140
gtaaagaaca tggatggact agcagaaata gacaccatga ttccatttga tctttcagcc	1200
accaagaaga acaccatggaa gatgtaccgt gtccaggctt cagataaaacc acacacagat	1260
gatccaaatc tctgcctcag tctttctcca gcctcgacc ctaggtctcg ccacaccatg	1320
ctaggagaga tattaaatatttta ttacacgcac tggccaggctt ccctgaagtt cacccctc	1380
ttctgtggctt ccatgtggc aactggaaaa ttatttagtat catatgcggcc gcccggggca	1440
gatccctccaa agaagaggaa ggaggccatg ctaggaactc atgtaatatg ggacattggg	1500
ttgcagagca gtcgcaccat ggttgtcccc tggatctcca acaccaccta caggcagacc	1560
attgtatgaca gcttcacaga aggtggttat atttctgtct tctaccagac aagaatttg	1620
gtgccgtgtt ccactccaag agaaatggat attcttggat ttgtatcagc ctgcaatgac	1680
ttctctgtga ggctgtgcg ggacaccact catattgagc agaaggcgt agcacaagga	1740
cttggacaaa tgctggagtc catgattgac aacactgtgc gggagactgt gggggcggcc	1800
acctcaagag atgctttggc aaatacagag gttcgggggc caacacacag caaagaata	1860
ccagcttga cagcagtaga aactggagca acaaattccac tagtcccttc agatactgt	1920
caaacaagac atgtggtgca gcacagatca agaagtggaa gttcaataga atccttctt	1980
gctcggtggg cctgtgtcac catcatgact gtggacaacc cagccctccac cacaacaaa	2040
gataaaattat ttgtgtgtg gaagatcacc tacaaagata ctgtcagct aagaagaaaa	2100
ctagagttct tcacctactc aagatttgac atggaactaa cttttgtggt cacagccac	2160
ttcacagaga ccaacaatgg acatgcttg aatcaagtctt accagattat gtacgtacca	2220
ccaggggctc ctgtgcccga aaaatggat gactacacgt ggcagacttc atcaaatcct	2280
tccatcttctt acacatatgg aacagcgccg gccaggatat cagttccctta tgtggccatc	2340
tcaaatgctt acagccactt ctatgtggc ttctccaaag taccttgaa ggaccagtcg	2400
ggggcgctgg gggacagttt atatggtgca gcatctctaa atgattttgg aatttggct	2460
gtccgtgttag taaatgacca caacccacc aaggtcacca gcaagatccg agtataattta	2520
aaacccaaagc acatccgagt gtgggttccg agaccgccaag gagctgttagc gtattacggt	2580
cctgggtgtgg actataaaga tggacttttg actcccttgc ccaccaaaaga cctcaccacg	2640
tac	2643

-continued

```
<210> SEQ ID NO 6
<211> LENGTH: 7441
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: synthetic
```

<400> SEQUENCE: 6

ttaaacacgc tctgggggtg taccacccccc agaggcccc acgtggccgta gtactccggat
attgcggta ccttgcacgc ctgtttata ctcccttccc gtaacttaga cgacaaaaac
caagttcaat agaagggggt acaaaccagt accaccacga acaagcactt ctgtttcccc
ggtgatgtcg tatagactgc ttgegtggtt gaaagcgacg gatccgttat ccgcttatgt
acttcgagaa gcccagtagc accctcgaaat ctgcgtcg ttgcgtcag cactcaaccc
cagagtgttag cttaggctga tgagtctgga catccctcac cgggtgacgggt ggtccaggct
gegttggcgcc cctacccatg gctaaccgca tgggacgcta gttgtgaaca aggtgtgaag
agcctattga gctacataaag aatccctccgg cccctgaatg cggctaatcc caaccctcgga
gcaggggtgatc acaaaccagt gattggcctg tcgtaacgcg caagttccgt gcggaaaccga
ctactttggg tgtccgtgtt tccttttatt ttattgtggc tgcttatgtt gacaatcaca
gattgttatac ataaagcgaa ttggattggc catccggtaa aagttagact cattatctat
ctgtttgcgt gatccgcgtcc attgagtggtt ttactctaa gtacaatttc aacagtttatt
tcaatcagac aattgtatca taatggcgcg gcaagtgatcg tctcaaaaagg ttggagccca
cgagaatagt aacagagcat acggatcaag cacaataat tatacgacga tcaattacta
ccgtgactcg gcatccaacg ccgctagcaa acaggacttt tctcaggatc catctaaatt
cacggaaaccc ataaaagacg tgcttattaa gaccgcacat atgctgaatt cacggatata
cgaagcctgt ggttactctg accgggtgct acagctgaca ctggcaact caacgatcac
aaccctaagag gcccggaaattt ccgttagtgc gtacggtagg tggccagagt acctactgt
ttcagaggctt aaccctgttag atcaacctac cgaaccggac gtggcagctt gtcgtttctta
cactcttgat accgtgtcct ggaccaggaa aagttaggggtt tgggtggtaa aactccccgaa
cgcccttagg gatatggggc tggtccgtca gaacatgtac tatcactatc tgggacgctc
aggctatacg gtacacgtgc aatgcacgc aagtaagttt catcaggggcg cgctggagtt
gttcgcagttt ccggaaatgtt gccttagctgg cgactcaaatac acaacgacta tgcatacaag
ttatcagaac gcaaaacccag gagaaaaggg gggAACATTACAGGACCATTTACCCAGA
taataaccaa acctccccag cacggagatt ctgtccgtta gactacttgtt tggggAACGG
aaccctactt ggtaacgcacat tcgttttcc acatcaaattt atcaatctta gaactataaa
ttgtcaactt ctgcgtttgc catatgtcaa ttccctgtcg atagactcca tggtaaaca
taacaactgg ggcatacgcca ttctaccgtt ccgcaccactt aacttcgcacat ccgaatccctc
acccgaaatc cccataaacac tcacaatgcg accaatgtgt tgcgaattta acggactgctc
gaatataacc ctaccggagac tccagggtttt acccgtaatg aatacgccag gatcaaattca
gtacttaacc gcccataattt tccaaatcccccc atgcgcctta ccagaatttcg acgttaacgc
tccaaattgtt attcggagac aggtgaaaaa tatgtatgggat ctcgcagaga ttgtataactt
gataccgttc gatctttcccg ctaccaagaa aaatacgatg gagatgttacc gctgtgagact
gtccgataag ccacacacccg atgacccaaat actgtgttttta agtttgctc cagcttccgaa
ccccacgtctc tctcataacca tqtttaqqqqa qataacttaac tattataactc attqqqccqq
2100

-continued

gtcattaaag tttacattcc tggggcggtt gtcatgtatgc gcccacggga agctgctagt	2160
tagctatgcc ccacctggtg cggtatccatc taaaagaga aaagaagcaa tgctcggtac	2220
ccacgttaatt tggatattt gactgcaatc gagctgcacc atgggtggtc cttggatatc	2280
taataacaacc tacagacaga caatcgacga ctcattcaca gagggtgggt atatatcggt	2340
gttttatcag acccgaatcg ttgtgccatt gtccacccct cggaaatgg atatccttgg	2400
cttcgtgatgt gcttgcacg actttcagt gagactacta agggacacta cgcatattga	2460
acagaaggca ctggctcaag gcctggcca aatgctcgag tcaatgtatcg ataatacgt	2520
gagagagacc gtaggtgcag cgacctcaag ggacgcgtt cctaataccg aggcgccgg	2580
gecgactcac agtaaagaga taccagcggtt aacagcagtc gaaaccgggtg cgactaatcc	2640
cttgggtccca agcgatacag tgcaacaag gcacgtatgc cagcacccat ccagggtcgga	2700
aagctctatc gagtcctttt ttgctagggg ggcttgcgtc actataatga cagtcgacaa	2760
tccggcatca acaaccaca aagacaagtt attcgctgtt tggaaaattt cctacaagga	2820
caccgtccaa ctgagacgga agttagaatt ctttacatac tcacggttcg acatggatt	2880
gactttcgttta gtgaccgcta actttaccga aaccaataac ggtcacgc tgaatcaggt	2940
ataccagatc atgtacgtcc ctccctgggc accagttccg gagaaatggg atgactatac	3000
ttggcagaca tctagcaacc cttagctttt ctacacttat ggtacccccc ccgctagaat	3060
tagtgtgcct tacgttagaa ttccaaacgc atataggccat ttctacgacg gattctcaaa	3120
gggtgcattt aaggatcaga gcgcggcctt gggggattcg ttgtacggag ccgcttccatt	3180
gaacgatttc gggatactag cagtcagggt cgtgaacgac cataacccta ctaaagtgac	3240
atcgaagata cgcgttacc ttaagccaaa acacatcagg gtgtgggtc ctggccgcc	3300
aagagccgtc gcctattatg gtcccggtt cgattataaa gacggAACCC tgacccatt	3360
gtcaactaaa gacctaacc cttatggatt cggacaccaa aacaaagccg tgtacactgc	3420
agtttacaaa atttcaact accacttggc cactcaggat gatttgcacaa acgcgtgaa	3480
cgtcatgtgg agtagagacc tcttagtca agaatcaaga gcccaggggca ccgattcaat	3540
cgcaagggtgc aatttcaacg caggggtgtt ctactgcgag tctagaagga aataactaccc	3600
agtatccctt gttggccaa cgttccagta catggaggct aataactatt acccagctag	3660
gtaccaggatcc catatgtca ttggccatgg attcgcatct ccaggggatt gtgggtgc	3720
actcagatgt caccacgggg ttagatggat cattactgtt ggtggagaag ggtgggtgc	3780
attttcagac attagagact tttatgccta cgaagaagaa gccatggaaac aaggcctcac	3840
caattacata gagtcacttgc gggccgcatt tggaaatggaa ttactcagc agattagcga	3900
caaaaataaca gagttgacca atatggtac cagtaccatc actgaaaagc tacttaagaa	3960
cttgcatacg atcatatcct cactagttat tataactagg aactatgaag acaccacaac	4020
agtgcgtcgct accctggccc ttcttgggtt tgatgcgttca ccatggcagt ggcttagaaa	4080
gaaagcatgc gatgttctgg agataccctt tgcgtatcaag caaggtgaca gttgggtgaa	4140
gaagtttact gaagcatgca acgcgttacg gggactggag tgggtgtcaa aaaaaatctc	4200
aaaatttattt gattggctca aggaaaaat tatccccacaa gctagagata agttggatt	4260
tgtacaaaa ctttagacaac tagaaatgtt ggaaaaccaa atctcaacta tacaccaatc	4320
atgccttagt caggaacacc agggaaattctt attcaataat gtcagatggt tatccatcca	4380
gtctaaagggtt tttggccctc ttacgcgtt ggaagccaaa agaatacaga aactcgagca	4440

-continued

tactattaac aactacatac agttcaagag caaacaccgt attgaaccag tatgtttct	4500
agtacatggc agcccccggaa caggtaaatc tgttagcaacc aacctgattt cttaggcatt	4560
agctgaaaga gaaaacacgt ccacgtactc gctaccccg gatccatcac acttcgacgg	4620
atacaaaca caggggatgg tgattatgga cgacctgaat caaaacccag atggtgccga	4680
catgaagctg ttctgtcaga tggtatcaac agtggagttt ataccaccac tggcatccct	4740
ggaggagaaa ggaatcctgt ttacttcaaa ttacgatctt gcatccacaa actcaagcag	4800
aatttcccccc cccactgtgg cacacagtga cgcgttagcc aggccgtttt cgttcgacat	4860
ggacatttcg gtcataatg agtattctag agatggaaaa ttgaacatgg ccatggctac	4920
tgaaatgtgt aagaactgtc accaaccacg aaactttaag agatgtgtc cttagtgtg	4980
tggtaaggca attcaattaa tggacaaatc ttccagatg agatacagta ttgaccagat	5040
cactacaatg attatcaatg agagaaacag aagatccaac attggcaatt gtatggaggc	5100
tttgttcaa ggaccactcc agtataaaga cttgaaaattt gacatcaaga cgagtcccc	5160
tccctgaatgt atcaatgact tgctccaagc agttgactcc caggaggatgg gagattactg	5220
tgagaagaag ggttggatag ttaacatcac cagccaggtt caaacagaaaa ggaacatcaa	5280
cagggcaatg acaattctac aagcggtgac aaccttcgac gcagttggctg gagttgtcta	5340
tgtcatgtat aaactgtttt ctggacacca gggagcatac actggtttac caaacaaaaa	5400
accccaacgtg cccaccattc ggacagcaaa ggtacaagga ccagggttcg attacgcagt	5460
ggctatggct aaaagaaaaca ttgttacagc aactactagc aaggggagat tcactatgtt	5520
aggagttccac gacaacgtgg ctatttacc aacccacgtc tcacctggtg aaagcatgt	5580
gatcgatggc aaagaagtgg agatcttgg tgccaaagcg ctcgaagatc aagcaggaac	5640
caatcttcaa atcactataa tcactctaa gagaatgaa aagttcagag acattagacc	5700
acatataacct actcaaatac ctgagacaaa tggatggatc ttgatgtca acactagca	5760
gtaccccaat atgtatgttc ctgtcggtc tggactgaa cagggtatc taaatctcg	5820
tgggcgc当地 actgtcgta ctctaattgtca caactttcca accagagcag gacagtgtgg	5880
tggagtc当地 acatgtactg ggaaaatgtcat cgggatgtcat gttgggggaa acgggttcaca	5940
cgggtttgca gcggccctga agcgatcata ctgcactcg agtcaagggtt aaatccagtg	6000
gatgagacct tcgaaggaag tgggatatcc aatcataat gccccgtcca aaaccaagct	6060
tgaacccagt gctttccact atgtgtttga aggggtgaag gaaccagcag tcctcactaa	6120
aaacgatccc aggtttaaga cagactttga ggaggcaattt ttctccaatg acgtggtaa	6180
caaaattact gaagtggatg agtacatgaa agaggcagta gaccactatg ctggccagct	6240
catgtcacta gacatcaaca tagacaatg tgcttggag gatgccatgt atggcactga	6300
tggcttagaa gcacttgatt tggatggatc ctttatgttag caatggaaaa	6360
gaagaagaga gacatcttga acaaacaac cagagacact aaggaaatgc aaaaactgct	6420
cgacacatg gaaatcaacc tcccactggt gacttatgtt aaggatgaac tttagatcaa	6480
aacaaagggtt gaggcaggaa aatccagatt aattgaatg tctagtttga atgactcagt	6540
ggcaatgaga atggcttttgg ggaacctata tgctgctttt cacaacacc cagggtgtat	6600
aacaggttca gcagtggtt gcgatccaga tttgttttgg agcaaaatcc cggtattgtat	6660
ggaaagagaag ctgtttgtt ttgactacac agggatgtat gcatctctca gcccgtctt	6720
gttcgaggca ctaaagatgg tgcttgagaa aatcgatggc ggagacagag ttgactacat	6780
cgactaccta aaccactcac accacctgttta caagaataaa acatactgtt tcaagggcg	6840

-continued

tatgccatct ggctgctcag gcacttcaat ttttaactca atgattaaca acttgattat	6900
caggacactc ttactgaaaa cctacaaggg catagattt gaccacctaa aaatgattgc	6960
ctatgggtat gatgttattt ctccctaccc ccatgaagtt gacgctagtc tccttagccca	7020
atcaggaaaa gactatggac taactatgac tccagctgac aaatcagctt catttggaaac	7080
agtcacatgg gagaatgtaa cattttgaa gagattttc agggcagacg agaaaataccc	7140
atttcttattt catccagtaa tgccaatgaa gggaaattcat gaatcaattt gatggacaaa	7200
agatccttagg aacactcagg atcacgttcg ctctctgtgc ctttttagctt ggcacaatgg	7260
cgaagaagaa tataacaaat tccttagctaa aatcaggagt gtgccaattt gaagagcttt	7320
attgctccca gagtactcaa cattgttaccg ccgttgggtt gactcatttt agtaacccta	7380
cctcagtcga attggattgg gtcatactgg tgttagggta aatttttctt taattcggag	7440
g	7441

<210> SEQ ID NO 7	
<211> LENGTH: 32	
<212> TYPE: DNA	
<213> ORGANISM: Unknown	
<220> FEATURE:	
<223> OTHER INFORMATION: synthetic	
<400> SEQUENCE: 7	
gcactgataa tgaactcctc tggatctact gg	32

<210> SEQ ID NO 8	
<211> LENGTH: 32	
<212> TYPE: DNA	
<213> ORGANISM: Unknown	
<220> FEATURE:	
<223> OTHER INFORMATION: synthetic	
<400> SEQUENCE: 8	
ccagtagatc cagaggagtt cattatcgt gc	32

<210> SEQ ID NO 9	
<211> LENGTH: 50	
<212> TYPE: DNA	
<213> ORGANISM: Unknown	
<220> FEATURE:	
<223> OTHER INFORMATION: synthetic	
<400> SEQUENCE: 9	
caagaattcc tgaccacata cggtgctcag gtttcatcac agaaagtgg	50

<210> SEQ ID NO 10	
<211> LENGTH: 50	
<212> TYPE: DNA	
<213> ORGANISM: Unknown	
<220> FEATURE:	
<223> OTHER INFORMATION: synthetic	
<400> SEQUENCE: 10	
caagaattcc tgaccacata cggtgctcag gtttcatcac agaaagtgg	50

<210> SEQ ID NO 11	
<211> LENGTH: 40	
<212> TYPE: DNA	
<213> ORGANISM: Unknown	
<220> FEATURE:	
<223> OTHER INFORMATION: synthetic	

-continued

aaagattaca ggtacacgta ccgatgccat agaggtgaca cacaaataca aaccgcaga	1740
tcatggaaa taaagaaact gtgggagcaa acccggttcca aagctggact gctggctcc	1800
gacggaggcc caaatttata caacattaga aatctccaca ttccctgaagt ctgcctaaaa	1860
tgggaattga tggatgagga ttaccagggg cgtttatgca acccactgaa cccatttgc	1920
agccataaaag aaattgaatc aatgaacaat gcagtgtatga tgccagcaca tggccagcc	1980
aaaaacatgg agtatgtatgc tggttgcaca acacactcct ggatccccaa aagaaatcg	2040
tccatcttga atacaagtca aagaggagta cttgaagatg aacaatgtatcc ccaaagggtgc	2100
tgcaatttat ttgaaaaatt ctccccccgc agttcataca gaagaccagt cgggatatcc	2160
agtatggtgg aggctatggt ttccagagcc cgaattgtatgc cacggattga ttgcataat	2220
ggaaggataa agaaagaaga gttcaactgatc atcatgaaga tctgttccac cattgaagag	2280
ctcagacggc aaaaatagtatcc aatttagctt gtccttcatg aaaaatgcc ttgtttctac	2340
t	2341

<210> SEQ ID NO 14

<211> LENGTH: 2341

<212> TYPE: DNA

<213> ORGANISM: unknown

<220> FEATURE:

<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 14

agcgaaagca ggcaaaccat ttgaatggat gtcaatccga ctttactttt cttaaaatgt	60
ccagcacaaa atgctataag cacaacttcc ctttatactg gagaccttcc ttacagccat	120
gggacacggaa caggatacac catggatact gtcaacaggaa cacatcgat ctcagaaaag	180
ggaagatgga caacaaacac cgaaactgga gcaccgcac tcaacccgat tggatggcca	240
ctgccagaag acaatgaacc aagtggttat gcccacacg atttgttattt ggaagcaatg	300
gttttccctt aggaatccca tcctggatttttggaaaact cgtgttatttga aacgtggag	360
gttgttcaggc aaacacgagt agacaagctg acacaaggcc gacagaccta tgactggact	420
ctaaatagaa accaacctgc tgcaacagca ttggccaaaca caatagaatgt gttcagatca	480
aatggctca cggccaatga gtctggaaagg ctcataacttc tccttaaggatcgatggag	540
tctatggaaa aagaggaaat ggggattacg acacatttc aacgaaaaag acgggtttagg	600
gataatatga caaaaaaaaaat gattacgaa cgaacaatcg gaaagaaaaa acagagactg	660
aataagcgat catacttgcat tagggcactt acacttaaca ctatgactaa ggacgcccggaa	720
aggggaaagc taaagcgtag agcaatttgc acacccggaa tgcaaatttgc ggggttgcgtat	780
tacttcgtcg agacactcgc tagatccata tgcgaaaatgt tagagcaatc cggactgcca	840
gtcgaaaaaaa acgaaaaactc gctaacgtcg ttggaaaat gatgactaat	900
agtcaggata cccgaaactgtc attacgatt accggcgata atactaagtgc gaacgagaat	960
cagaatccata gaatgtttct cgcaatgtatc acatatatgtatc cacgtacca acccgaatgg	1020
tttagaaacg tactgtcaat cgcaccaattt atgttttagca ataagatggc tagattggc	1080
aaggggtata tgtttgcata taaggtatgc aaatttgcgaa cacagatacc tgccgaaatg	1140
ctagcatcaa tcgatctaaa gtactttaac gatagtgatc gaaaaaaaat cgaaaagat	1200
agaccgttac tgatagaggg aaccgcgcgc cttatccccgc gaatgtatgc ggggttgcgtat	1260
aatatgctta gtaccgttgc aggcgttgc atacttaact tagggcaaaa acgttataact	1320

-continued

aagactacat	attgggtggga	cggactgcaa	tctagcgacg	atttcgact	aatcgtaac	1380
gcacctaacc	atgaggggat	acaaggccga	gtcgatagat	tctatagaac	atgcaaaactg	1440
ttagggatta	atatgtctaa	aaaaaagtca	tacataaata	gaaccggaac	atttgaattc	1500
actagtttt	tttacagata	cggattcggt	gctaatttt	gtatggagtt	acctagttc	1560
ggagtagcg	gaattaacga	atccgcccgt	atgtcaatcg	gctgttaaccgt	tattaagaat	1620
aatatgatta	ataacgatct	agggccagea	accgcacaaa	tggcattgca	gttgttcata	1680
aaggattatac	gttatacata	tagatgtcat	agaggcgata	cacagataca	gactagacga	1740
tcatttggaa	tcaaaaaatt	gtgggagcaa	actaggctca	aagccggact	gttagtgtcc	1800
gacggagggc	ctaattctata	caatattagg	aatctgcata	tacccgaagt	gtgtctaaag	1860
tgggagctta	tggacgaaga	ctatcagggg	agattgtgca	atccgcttaa	cccattcggt	1920
agccataaaag	agatagagtc	aatgaataac	gccgttatga	tgccagcaca	cggaacccgct	1980
aagaatatgg	aatacgacgc	agtcgcaact	acacatagtt	ggataccgaa	acggaatcga	2040
tccatactga	atacatcccc	aagaggcgta	ctcgaagacg	aacaaatgta	ccaacgggt	2100
tgcaatctat	ttgaaaaatt	ttttcctagt	agtagctata	gacgaccagt	cgggatatcc	2160
agtatggtgg	aggctatggt	ttccagagcc	cgaattgtat	cacggattga	tttcgaatct	2220
ggaaggataa	agaaagaaga	gttcaactgag	atcatgaaga	tctgttccac	cattgaagag	2280
ctcagacggc	aaaaatagt	aatttagctt	gtccttcatg	aaaaaatgcc	ttgtttctac	2340
t						2341

<210> SEQ_ID NO 15
 <211> LENGTH: 2341
 <212> TYPE: DNA
 <213> ORGANISM: unknown
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 15

agcgaaagca	ggcaaaccat	ttgaatggat	gtcaatccga	ccttactttt	cttaaaagt	60
ccagcacaaa	atgctataag	cacaactttc	ccttatactg	gagaccctcc	ttacagccat	120
ggcacagggaa	caggatacac	catggatact	gtcaacagga	cacatcagta	ctcagaaaag	180
ggaagatggaa	caacaaacac	cggaaactgga	gcaccgcaac	tcaaccggat	tgatggcca	240
ctgccagaag	acaatgaacc	aagtggttat	gcccaaacag	attgtgtatt	ggaagcaatg	300
gtttcccttg	aggaatcccc	tcctggattt	tttgaaaact	cgtgttattga	aacgatggag	360
gttggtcagc	aaacacggat	agacaagctg	acacaaggcc	gacagaccta	tgactggact	420
ctaaatagaa	accaacctgc	tgcaacagca	ttggccaaca	caatagaagt	gttcagatca	480
aatggcctca	oggccaatga	gtctggagg	ctcatagact	tccttaagga	cgttatggag	540
tctatggaaa	aagaggaaat	ggggattacg	acacatttc	aacgaaaaag	acgggtttagg	600
gataatatga	aaaaaaaaat	gattacgca	cgaacaatcg	gaaagaaaaa	acagagactg	660
aataagcgat	catacttgat	tagggactt	acacttaaca	ctatgactaa	ggacgcccga	720
aggggaaagc	taaagcgtag	agcaattgca	acacccggaa	tgcaaattag	ggggttcgta	780
tacttcgtcg	agacactcgc	tagatccata	tgcgaaaagt	tagagcaatc	cggaactgcca	840
gtcgaaaaaaa	acgaaaaaaa	agcgaaactc	gctaacgtcg	ttagaaaaat	gatgactaat	900
agtcaggata	ccgaactgtc	atttacgatt	accggcgata	atactaagt	gaacgagaat	960
cagaatccta	aatgtttct	cgcaatgatc	acatatatga	cacgtaacca	acccgaatgg	1020

-continued

ttagaaacg tactgtcaat cgcaccaatt atgtttagca ataagatggc tagattggc	1080
aagggtata tgtttatac taagagtatg aaattgcgaa cacagatacc tgccgaaatg	1140
ctagcatcaa tcgatctaaa gtacttaac gatagtagac gaaaaaaaaat cgaaaagatt	1200
agaccgttac tgatagaggg aaccgccagc ctatccccg gaatgtatgat ggggatgttt	1260
aatatgttta gtaccgttta aggcgttagc atacttaact tagggcaaaa acgttataact	1320
aagactacat attggtgaaa cggactgca tctagcgacg atttcgact aatcgtaac	1380
gcacctaacc atgaggggat acaagccgga gtcgatagat tctatagaac atgcaactg	1440
ttagggatta atatgtctaa aaaaaagtca tacataaata gaaccggaaat ttgttattc	1500
acaagtttt tctatcgta tgggtttgtt gccaatttca gcatggagct ccccagttt	1560
ggggtgtctg ggatcaacga gtcagcgac atgagtattt gagttactgt cataaaaaac	1620
aatatgataa acaatgtatct tggccagca acagctcaaa tggcccttca gttgttcatc	1680
aaagattaca ggtacacgta ccgatgccat agaggtgaca cacaataaca aacccgaaga	1740
tcatttggaaa taaagaaact gtgggagcaa acccggttcca aagctggact gctggctcc	1800
gacggaggcc caaatttata caacatttata aatctccaca ttccctgaatg ctgcctaaaa	1860
tgggaattga tggatgagga ttaccagggg cgtttatgca acccactgaa cccatttgc	1920
agccataaaag aaattgtatc aatgtacaaat gcagtgtatg tgccagcaca tggccagcc	1980
aaaaacatgg agtatgtatc tggtaacaaca acacacttcc ggtatcccaa aagaaatcga	2040
tccatcttgc atacaagtca aagaggagta cttgaagatg aacaaatgtt ccaaagggtc	2100
tgcaattttat ttgaaaaatt ctcccccagc agttcataca gaagaccagt cgggatatcc	2160
agtatgggtgg aggctatggt ttccagagcc cgaattgtatg cacggattga ttctgaatct	2220
ggaaggataa agaaaagaaga gttcaactgatc atcatgaaga tctgttccac cattgaagag	2280
ctcagacggc aaaaatagtg aatttagtt gtccttcatg aaaaaatgcc ttgtttctac	2340
t	2341

<210> SEQ ID NO 16

<211> LENGTH: 2341

<212> TYPE: DNA

<213> ORGANISM: Influenza virus

<400> SEQUENCE: 16

agcgaaagca ggtcaattat attcaatgt gaaagaataa aagaactaag aaatctaatt	60
tccgagtctc gcacccgcga gatactcaca aaaaccaccc tggaccatat ggcataatc	120
aagaagtaca catcaggaag acaggagaag aacccagcac ttaggtatgaa atggatgt	180
gcaatgaaat atccaattac agcagacaag aggataacgg aaatgttcc tgagagaaat	240
gagcaaggac aaactttatg gatgttccatg aatgtatgcg gatcagaccc agtgtatgt	300
tccacccatgg ctgtgacatg gtggatagg aatggaccaa taacaaatac agttcattat	360
ccaaaaatct acaaaaactta ttttggaaaga gtcgaaaggc taaagcatgg aacctttggc	420
cctgtccatt ttagaaacca agtcaaaata cgtcgagag ttgacataaa tcctggcat	480
gcagatctca gtgccaagga ggcacaggat gtaatcatgg aagttgttt ccctaacgaa	540
gtggggccca ggataactaac atcggaatcg caactaacga taaccaaaga gaagaaagaa	600
gaactccagg attgcaaaat ttctcccttg atgggttgc acatgttgg aagagaactg	660
gtccgcaaaa cggatccct cccagtggtt ggtggaaaca gcaatgtgtt cattgaagtg	720

-continued

ttgcatttga	ctcaaggaac	atgctggaa	cagatgtata	ctccaggagg	ggaagtgagg	780
aatgatgatg	ttgatcaaag	cttgattatt	gctgcttaga	acatagttag	aagagctgca	840
gtatcagcag	atccactagc	atcttattt	gagatgtgcc	acagcacaca	gattgggtgga	900
attaggatgg	tagacatcct	tagcagaac	ccaacagaag	agcaagccgt	ggatatatgc	960
aaggctgcaa	tgggacttag	aattagctca	tccttcagtt	ttgggtggatt	cacatthaag	1020
agaacaagcg	gatcatcagt	caagagagag	gaagaggtagc	ttacggggca	tcttcaaaca	1080
ttgaagataa	gagtgcatga	gggatatgaa	gagttcacaa	tggggggag	aagagcaaca	1140
gcctactca	aaaaagcaac	caggagattt	attcagctga	tagtgagtgg	gagagacgaa	1200
cagtcgattt	ccgaagcaat	aattgtggcc	atggtatattt	cacaagagga	ttgtatgata	1260
aaagcgtca	gagggtatct	gaattcgtc	aataggcgca	atcagcgatt	gaatcctatg	1320
catcaactt	taagacattt	tcagaaggat	gcgaaagtgc	ttttcaaaa	ttggggagtt	1380
gaacctatcg	acaatgtat	gggatattgc	cagacatgac	tccaaagcatc		1440
gagatgtcaa	tgagaggagt	gagaatcgc	aaaatgggt	tagatgatgc	ctccagcagc	1500
gagagggttag	tggtgagcat	tgaccgtttt	ttgagaatcc	gggaccaacg	aggaaatgt	1560
ctactgtctc	ccgaggaggt	cagtgaaaca	caggaaacag	agaaactgac	aataacttac	1620
tcatcgtaa	tgtgtggg	gattaatgg	cctgaatcag	tgttgttcaa	tacctatcaa	1680
tggatcatca	gaaactggg	aactgttaaa	attcagttgt	cccagaaccc	tacaatgtca	1740
tacaataaaa	tgaaatttga	accatttcag	tctttatgtac	ctaaggccat	tagaggccaa	1800
tacagtgggt	ttgttaagaac	tctgttccaa	caaattgggg	atgtgttgg	gacatttgat	1860
accgcacaga	taataaaaact	tcttcccttc	gcagccgctc	caccaaagca	aagtagaaatg	1920
cagttctct	catttactgt	gaatgtgagg	ggatcaggaa	tgagaataact	tgttaaggggc	1980
aattctctg	tattcaacta	taacaaggcc	acgaagagac	tcacagttct	cgaaaaggat	2040
gctggcacct	taactgaaga	cccagatgaa	ggcacagctg	gagtgaggct	cgctgttctg	2100
aggggattcc	tcattctggg	caaagaagac	aagagatatg	ggccagcaat	aagcatcaat	2160
gaactgagca	accttgcgaa	aggagagaag	gctaattgtgc	taattgggca	aggagacgt	2220
gtgttgtaa	tgaaacggaa	acgggactct	agcatactta	ctgacagcc	gacagcgacc	2280
aaaagaattc	ggatggccat	caattatgt	cgaatagttt	aaaaacgacc	ttgtttctac	2340
t						2341

<210> SEQ ID NO 17

<211> LENGTH: 2341

<212> TYPE: DNA

<213> ORGANISM: unknown

<220> FEATURE:

<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 17

agcgaaagca	ggtcaattat	attcaatatg	gagagaatca	aagagcttag	gaatctttag	60
tcacaatcta	gaactagaga	gatactgact	aagactacag	tcgatcatat	ggctataatc	120
aaaaaatata	ctagcggaaag	acaggaaaaaa	aatcccgac	ttagaatgaa	atggatgtat	180
gctatgaaat	accctattac	agccgataag	cgaatttaccg	aatgatacc	agagagaaac	240
gaacaggagc	agacattgtg	gtctaaaatg	aacgacgccc	gatccgatag	agtgtatgg	300
tgcgccactag	ccgtaacatg	gtggaaataga	aacggaccta	ttacgaatac	agtgcattac	360
cctaagatat	acaaaacata	tttcgaaaga	gtcgagagac	tgaaacacgg	aacattcgga	420

-continued

ccagtgcatt ttccggaaatca ggttaagatt agacgttagag tgcgatattaa tccaggggcat
gcagatctc ccgctaaaga ggcacaagac gttattatgg aggtcggtt tcctaaccgag
gtcgccgctt ggataactgac tagcgaatcg caattgacaa ttacgaaaaga gaaaaaaagag
gaactccagg attgcaaaat tagcccactt atggtcgcat atatgctcgac acgcgaattt
gttagaaaga ctagattcct accagtcgca ggccggaaacgt ctagcgtgttatcgaagtg
ttgcatacata cacaggaaac atgttggag caaatgtata ctccaggagg cgaagtggaga
aacgacgacg ttgatcaatc gctaattcata gccgcttagga atatagtgag aaggcgagcc
gttagcgcacg accccacttgc gtcactactc gaaatgtgcc atagtcgcac aatcgaggg
attagaatgg tcgatatacct taggcagaat cctacagagg aacaggccgt agacatatac
aaagccgcaa tgggattgcg aattagctca tcattctcat tcggagggtt tacgtttaaa
cggaactagcg gatctagcgt aaaacgcgaa gaggaagtgc ttactggcaa tctgc当地
ctaaagatta ggggtgcata gggatacga gagtttacaa tggtcgacg tagagcaacc
gtctatactta gaaaacgcac taggagactg atacaattga tcgttagcg aaggacgaa
cagtcaatcg ccgaagcgat aatagtcgca atgggtttt cgcaagagga ttgcattt
aaggccgtta ggggggatct gaatttcgtt aataggcata atcagagact gaatccat
catcaattgc ttagacatct tcagaaagac gctaaatgt tgtttacaa ttggggatc
gaacctatcg ataacgttat gggatgtata gggatactgc cagatatgac accatcaatc
gaaatgtcaa tgagaggcgt taggattgt aagatggcgt tagacgata ctccagact
gagagagtgg tagtgtcaat cgatagatit cttaggatta gggatcagag aggcaacgt
ctgtctatcac ccgaagaagt tagcgaacaa cagggAACCG aaaaatttgc aattacgtat
agtagtagta tgatgtggg gattaacgga ccagactcg tgtagtgaa tacatata
tggataatac ggaattggg gacagtggaa atacaatgtt cacagaatcc tacaatgt
tacaataaga tggagttcga acctttcaa tcgttagtgc ctaaggccat aagaggccaa
tataatggg tgcgttacaaat ttgtttcgtt caaatggatcg acgtactcg aacattcgat
accgcacaga taattaagct attgcatttc gcagccgcac cacctaagca atctagaatg
caatttctt gtttaccgt taacgttagg ggttccggaa tgcaataact cgtaggggg
aatagttccat tggttacaaat caataaggca actaagagat tgacagtgtt aggcaaggac
gcagggacat tgaccgaaga cccagacgag ggaaccgcgtg ggttccggaa tgcaataact
agggggtttc tgatgtccgg aaaggaggat aagagatcg gacgtcgact atcgatata
gaactatcta atctcgctaa aggcgaaaaa gcaatgtgtt taatcgacca gggagacgta
gtgttagtgc tgaaacggaa acgcgatagc tcaataactga cagactcaca aaccgcact
aagagaattt ggttccggaaat taatttagtgc cgaatgtttt aaaaacgcacc ttgtttctac
t

<210> SEQ ID NO 18
<211> LENGTH: 2233
<212> TYPE: DNA
<213> ORGANISM: Influenza virus

<400> SEQUENCE: 18

```
agcgaaagca ggtactgatc caaaaatggaa gatTTTGTGC gacaatgctt caatccgatg 60  
atTTGTCGAGC ttgcggaaaa aacaatgaaa qagtatgggg aggacCTGAA aatcgaaaca 120
```

141

142

-continued

```
<210> SEQ ID NO 19
<211> LENGTH: 2233
<212> TYPE: DNA
<213> ORGANISM: unknown
<220> FEATURE:
<223> OTHER INFORMATION: synthetic
```

-continued

<400> SEQUENCE: 19

agcgaaagca ggtactgatc caaaaatggag gatttcgtta ggcaatgcct taatccatg	60
atagtcgagt tagccgaaaa gactatgaaa gagtatggcg aagacctaaa gattgagact	120
aataaaattcg ccgcaatttg cacacacccctt gaggtttgtt ttatgtatcc cgatcccac	180
tttattaacg aacagggaga gtcaattata gtcgagtttag gcgatccgaa cgcattgcta	240
aagcatagat ttgaaattat agaggggacgc gataggacaa tggcatggac cgtatgttata	300
tcgatggca atacaaccgg agccgaaaaa ccgaaattctt taccggatct atacgattat	360
aaagagaata ggtttatcga aatcgagtg actagacgcg aagtgcataat ttattatctc	420
gaaaaagcga ataagattaa gtccgaaaag acacacatac acattttag cttaaccgga	480
gaggaaatgg caacaaaagc cgattataca ctggacgaaag agtcttagggc taggattaag	540
actagactgt ttacaattatg acaggaaatg gcttagtaggg ggttgtggg tagctttaga	600
caatccgaaa gaggcgaaga gacaatcgaa gagagatttgg aaattaccgg aacaatgcga	660
aagcttgccg atcaatccctt acccccccaat ttctctagcc ttgagaattt tagggcatac	720
gttgacggat tcgaacctaa cggttatata gaggaaagc tatcgcaaat gtctaaagag	780
gttaacgcta gaatcgaaacc attcttaaag acaacaccta gaccacttag actgccaac	840
ggaccaccat gtcacagcg atctaagttt ctgcttatgg acgcactaaa gttgtcaatc	900
gaagaccat cacaegaggg agaggggata ccattgtacg acgcaattaa gtgtatgcga	960
acattttcg gatggaaaga gcctaaccgtt gtggaaaccac acgaaaaagg gattaatccg	1020
aattatctgc ttagttggaa acaggtgttta gcccattttgc aggatatcgaa aacgaagag	1080
aaaattccgaa aactaagaa tatgaaaaaa actagccaaatc tgaaatgggc acttggcgag	1140
aatatggcac ccgaaaaagt cgatttcgac gattgcaaaatc acgtcgccgaa tctaaagcaa	1200
tacgatagcg acgaaccggaa acttagatca ctgcgtatgg ggatacagaa cgagttcaat	1260
aaggcatcg aattgaccga tagctcatgg atagagtttgc acgagatagg cgaagacgtt	1320
gcaccaatcg aacacatagc ctctatgaga cggattttt ttacatccgaa agtgcacat	1380
tgttagggcaa cagagtatat tatgaaaggg gtgtatattttt ataccgcatt gcttaacgct	1440
agttgegccc caatggacga ttccaaactg ataccgtatc tctcgaaatgg tagaaacaaa	1500
gagggacgta gaaagactaa tctgtatggg ttcatttttta agggaaaggc tcatttaagg	1560
aacgatacag acgttagtgaa ttccgtatggt atggagtttgc gccttaccgg tccgagactc	1620
gaaccacaca aatggaaaaa gtattgcgtt ctagagatgg gggatgtt gattagatcc	1680
gcaatcgac aggtttcgag accaatgtttt ttgtacgttta ggactaacgg aacctcgaa	1740
ataaaaatgaa aatggggat ggagatgcgtt agatgcctat tgcataatccct tcagcaatc	1800
gaatctatgaa tagaggccgaa atctcgatggt aaagagaaatc atatgacaaa aggtttttt	1860
gaaaataagt ccgaaacatg gccaatcgaa gagtcaccaa aaggggttga ggaatccctca	1920
atcgaaaaag tttgttagaaac attgctcgatggt aaatccgtat tcaatagtctt atacgcccagc	1980
ccacaactag agggattctc tgctgagttca cggaaactgt tactgtatgtt gcaagccctt	2040
agggataatc tcgaaccggaa aacattcgat ctaggggggtt tgcgtacggc aatcgaaagag	2100
tgtctgatggt acgatccatg ggtactgtt aacgcttagt ggtttatc gttccttaca	2160
cacgcactat ctttagttgtt gcaatcgatc tatttgcattt ccataactgtc caaaaaagta	2220
ccttggttctt act	2233

-continued

<210> SEQ ID NO 20
<211> LENGTH: 1775
<212> TYPE: DNA
<213> ORGANISM: Influenza virus

<400> SEQUENCE: 20

agcaaaaagca	ggggaaaata	aaaacaacca	aaatgaaggc	aacacctactg	gtcctgttaa	60
gtgcacttgc	agctgcagat	gcagacacaa	tatgtatagg	ctaccatgcg	aacaattcaa	120
ccgacactgt	tgacacagta	ctcgagaaga	atgtgacagt	gacacactct	gttaacctgc	180
tcgaagacag	ccacaacgga	aaactatgt	gattaaaagg	aatagccca	ctacaattgg	240
ggaaatgtaa	catcgccgga	tggctttgg	gaaacccaga	atgcgacc	ctgcttcag	300
tgagatcatg	gtcctacatt	gtgaaacac	caaactctgt	aatggataa	tgttatccag	360
gagatttcat	cgactatgag	gagctgaggg	agcaattttag	ctcagtgtca	tcattcgaaa	420
gattcgaaat	atttccaaa	gaaagctcat	ggcccaacca	caacacaaaac	ggagtaacgg	480
cagcatgctc	ccatgagggg	aaaaggcgtt	tttacagaaa	tttgctatgg	ctgacggaga	540
aggagggctc	atacccaaag	ctgaaaaatt	cttatgtgaa	caaaaaagg	aaagaagtcc	600
ttgtactgtg	gggtattcat	cacccgccta	acagtaagga	acaacagaat	atctatcaga	660
ataaaaatgc	ttatgtctct	gtagtgtactt	caaattataa	caggagattt	accccgaaaa	720
tagcagaaag	ccccaaagta	agagatcaag	ctggggaggat	gaactattac	tggaccttgc	780
taaaacccgg	agacacaata	atatttgagg	caaattggaaa	tctaatagca	ccaatgtatg	840
cttcgcact	gagtagaggc	tttgggtccg	gcatcatcac	ctcaaaccgca	tcaatgtcat	900
agtgtaacac	gaagtgtcaa	acacccctgg	gagctataaa	cagcagtctc	ccttaccaga	960
atatacaccc	agtccacaata	ggagagtgcc	caaatacgt	caggagtgcc	aaatttggaga	1020
tggttacagg	actaaggaac	actccgtcca	ttcaatccag	aggtctat	tttggccattt	1080
ccgggtttat	tgaaggggga	tggactggaa	tgtatagtt	atggatgtgt	tatcatcatc	1140
agaatgaaca	gggatcaggc	tatgcagcgg	atcaaaaaag	cacacaaaat	gccattaacg	1200
ggattacaaa	caaggtgaac	actgttatcg	agaaaatgaa	cattcaattc	acagctgtgg	1260
gtaaagaatt	caacaaatta	gaaaaaagg	tggaaaattt	aaataaaaaa	gttgtatgt	1320
gatttctgga	catttggaca	tataatgcag	aatttgttagt	tctactggaa	aatgaaagga	1380
ctctggattt	ccatgactca	aatgtgaaga	atctgtatga	gaaagtaaaa	agccaattaa	1440
agaataatgc	caaagaatac	ggaaatggat	gttttgagg	ctaccacaag	tgtgacaatg	1500
aatgcata	aaatgtgaaga	aatgggactt	atgattatcc	caaataattca	gaagagtcaa	1560
agttgaacag	ggaaaaggta	gatggagtga	aatttggaaatc	aatggggatc	tatcaggattc	1620
tggcgatcta	ctcaactgtc	gccagttcac	tggtgctttt	ggtctccctg	ggggcaatca	1680
gtttctggat	gtgttcta	atgtttttgc	agtgcaaat	atgcatactg	gattagaatt	1740
tcagaaatat	gaggaaaaac	acccttgttt	ctact			1775

<210> SEQ ID NO 21
<211> LENGTH: 1775
<212> TYPE: DNA
<213> ORGANISM: unknown
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 21

agcaaaaagca	ggggaaaata	aaaacaacca	aaatgaaggc	aacacctactg	gtcctgttaa	60
-------------	------------	------------	------------	-------------	------------	----

-continued

gtgcacttgc agctgcagat gcagacacaa tatgtatagg ctaccatgca aacaattcaa	120
ccgacactgt tgacacagta ctcgagaaga atgtgacagt gacacactt gttaacctgt	180
tagaggactc acataacgga aagctatgta ggcttaaggg aatcgacca ctgcaattgg	240
gcaagtgtaa tatagccgga tggttgttg ggaatccgga atgcgatcca ctgttacccg	300
ttaggtcatg gtcatatata gtcgagacac ctaatagcga aaacggaatt tgttatcccg	360
gcgattttat cgattacgaa gagcttagag agcaattgtc tagcgttagt tcattcgaaa	420
gatcgaaat tttccgaaa gagcttagtt ggccaaatca taatactaac ggagtgactg	480
ccgcatgctc acacgaaggc aagtcttagt tttagggaa tctgttggg ttgactgaga	540
aagagggatc atatccgaaa ctgaaaaact catacgtaa caaaaaggaa aaggaagtgt	600
tagtgttgc ggggatacac catccaccaa atagtaaaga gcaacagaat atatatacaga	660
acgaaaaacgc atacgtagc gtcgtaacta gtaattataa tagaaggaaa acacccgaaa	720
tgcagagag accgaaagtt agagaccaag ccgaaagaat gaattattat tggacactac	780
tgaaacccgg cgatacaatt atattcgaag cgaacggaaa tctgatcgca ccgatgtatg	840
cattcgaact atctaggggg ttccggatccg gaattattac tagtaacgct agtatgcacg	900
aatgtAACAC gaagtgtcag actccactag ggcgaattaa ctctagtcg ccatatcaga	960
atatacatcc cgtaacaatc ggcgaatgcc caaaatacg tagatccgct aagcttagaa	1020
tggttaccgg actgagaaat acaccatcaa tccaatctag ggggttggc ggagcgatag	1080
ccggatttat cgaagggggg tggacaggga tgatagacgg ttggtacgga tatcatcacc	1140
aaaacgaaca gggatccgga tacgcagccg atcagaaatc gacgcaaaac gctattaacg	1200
gaattactaa taaagtgaat accgtaatcg aaaaaatgaa tatccaattt accgcagtcg	1260
gaaaggaatt caataagctt gagaaaagaa tggagaatct gaataaaaaa gtgcacgacg	1320
gatttctaga catatggact tataacgccc aactgttagt gttgctcgaa aacgaaagaa	1380
cactagactt tcacgactca aacgtaaga atctatacgaaa aatgtggaaa tcccaattga	1440
aaaataacgc taaagagata gggAACGGAT gtttcgagtt ctatcataaa tgcgataacg	1500
aatgtatgga atccgtttagg aacgaaacat acgattatcc taagtatagc gaagagtcaa	1560
aactgaatag ggagaaagtc gacggagtga aactcgaatc aatggggata tatcagatac	1620
tggcaatcta tagtacagtc gccagctcac tggttttttt ggtctccctg gggcaatca	1680
gtttctggat gtgttctaat ggtctttgc agtgcagaat atgcatacg gattagaatt	1740
tcagaaatat gaggaaaaac acccttgttt ctact	1775

<210> SEQ ID NO 22

<211> LENGTH: 1565

<212> TYPE: DNA

<213> ORGANISM: Influenza virus

<400> SEQUENCE: 22

agcaaaaagca gggtagataa tcactcactg agtgacatca aaatcatggc gtcccaaggc	60
accaaaacggt ottacgaaca gatggagact gatggagaac ggcagaatgc cactgaaatc	120
agagcatccg tcggaaaaat gattgggtga attggacgt tctacatcca aatgtgcacc	180
gaactcaaac tcagtgatta tgaggacgg ttgatccaa acagcttaac aatagagaga	240
atggtgctct ctgctttga cggaaaggaga aataaaatacc tggagaaca tcccgagtgcg	300
gggaaagatc ctaagaaaaac tggaggacct atatacagga gagtaaacgg aaagtggatg	360

-continued

agagaactca tccttatga caaagaagaa ataaggcgaa tctggcgcca agctaataat	420
ggtgacgatg caacggctgg tctgactcac atgatgtatc ggcattccaa tttgaatgat	480
gcaacttatac agaggacaag agctcttggc cgacccggaa tggatcccag gatgtgcct	540
ctgatgcaag gttcaactct ccctaggagg tctggagccg cagggtgc agtcaaagga	600
gttggacaa tggtgatgga attggtcagg atgatcaaac gtgggatcaa tgatcgaaac	660
ttctgggggt gtgagaatgg acgaaaaaca agaattgctt atgaaagaat gtgcaacatt	720
ctcaaaggaa aatttcaaac tgctgcacaa aaagcatga tggatcaagt gagagagac	780
cggAACCCAG ggaatgctga gttcgaagat ctcactttc tagcacggtc tgcaactata	840
ttgaggggtt cggttgctca caagtcctgc ctgcctgcct gtgtgtatgg acctgcccgt	900
gccagtggtt acgactttga aagagaggaa tactctctag tcggaataga cccttcaga	960
ctgcttcaaa acagccaagt gtacagccta atcagaccaa atgagaatcc agcacacaag	1020
agtcaactgg tgtggatggc atgcattctt gccgcatttg aagatctaag agtattaagc	1080
ttccatcaaag ggacgaaggt gctcccaaga gggaaagctt ccactagagg agttcaatt	1140
gcttccaatgaaaatggaa gactatggaa tcaagttacat ttgaactgag aagcaggat	1200
tggccataa ggaccagaag tggaggaaac accaatcaac agagggcata tgcggggccaa	1260
atcagcatac aacctacgtt ctcagtagag agaaatctcc ctttgacag aacaaccatt	1320
atggcagcat tcaatggaa tacagaggaa agaacatctg acatgaggac cgaaaatcata	1380
aggatgtatgg aaagtgcag accagaagat gtgtcttcc aggggggggg agtctcgag	1440
ctctcgacg aaaaggcagc gagccccatc gtgccttc ttgacatgag taatgaagga	1500
tcttatttct tcggagacaa tgcagaggag tacgacaatt aaagaaaaat accctgttt	1560
ctact	1565

<210> SEQ_ID NO 23

<211> LENGTH: 1565

<212> TYPE: DNA

<213> ORGANISM: unknown

<220> FEATURE:

<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 23

agcaaaaagca gggtagataa tcactcactg agtgacatca aaatcatggc gtcccaaggc	60
acccaaacggt cttacgaaca gatggagact gatggagaac gccagaatgc cactgaatc	120
agagcttagcg tcggaaaaat gatagggggaa atcggaggt tttacataca aatgtgtacc	180
gaactcaaata tgcgtcgatca cgaaggggaa ttgatccaa atagtctgac aatcgaaaga	240
atggtgttaa gcgcattcga cggaaacacgg aataagtatc tgcggatggca tccttagcgca	300
ggcaaggatc caaaaaaaaac cggaggggca atctatagga gatgtggatc aaagtggatg	360
cgcgaactga tactgtacga taaagaggag attagacgggat tgcggatggca agcgaataac	420
ggagacgacg ctactgcccgg actgacacat atgatgtatc ggcactctaa tcttaacgc	480
gtcatacacc aacggacttag ggcactcggtt agaaccggaa tggatccatc aatgtgc	540
ctttagcgagg gatctacact cccttagacga tccggagccg caggagcagc cggtttaggg	600
gtcggacta tggatggaa actcggttgc atgataaaaa gggggatggaa cgatggat	660
ttttggagag gcgaaaacgg acgtaaaact agaatcgat acgaaagaat gtgcaatata	720
ctcaaaggaa aatttcaaac cgccggccaa aaagctatga tggatcaagt tagggatct	780
aggaatccag gaaatgcccga attcgaaac cttacatttc tgcgtcggtc cgccactata	840

-continued

cttcgcggat	cagtcgcaca	caaattctgc	ttaccgcat	gctgtatacg	acctgcgtc	900
gctagcggat	acgatttcga	acgcgaaggg	tatagtctag	taggaattga	tccattttaga	960
tttgtccaaa	atttcgaagt	gtatagtcgt	attagaccta	acgagaatcc	tgcacacaaa	1020
tctcaactcg	tatggatggc	atgccatagt	gcccgcattcg	aagaccttag	agtgcgtatct	1080
ttcataaaagg	gaacgaaagt	gttgcgttagg	ggaaagctat	ctactagggg	agtgc当地	1140
gcttagtaacg	agaatatgga	gactatggag	tcttagtacac	tcgaaactgag	atctagatata	1200
tgggctatta	ggactagatc	cgaggaaat	acgaatcgc	aacgagctag	cggccggcaaa	1260
atctcaatcc	aacctacatt	ttccgtgcaa	cggaatctgc	cattcgatcg	gacaacgatt	1320
atggccgcata	tcaatggaa	taccgaggga	cggaactagcg	atatgagaac	cgaaattatc	1380
agaatgatgg	aatccgctag	accagaggac	gttccgtttc	aaggacgggg	agtcttcgag	1440
ctctcggacg	aaaaggcagc	gagccccgatc	gtgccttcct	ttgacatgag	taatagaaggaa	1500
tcttatttct	tcggagacaa	tgcagaggag	tacgacaatt	aaagaaaaat	acccttgaaa	1560
ctact						1565

<210> SEQ ID NO 24
<211> LENGTH: 1413
<212> TYPE: DNA
<213> ORGANISM: Influenza virus

<400> SEQUENCE: 24

agcgaaagca ggggtttaaa atgaatccaa atcagaaaat aacaaccatt ggatcaatct 60
gtctggtagt cggactaatt agcctaataat tgcaaataagg gaatataatc tcaatatgg 120
ttagccattc aattcaaact ggaagtcaaa accatactgg aatatgcaac caaaacatca 180
ttacctataa aaatagcacc tggtaaagg acacaacttc agtgatatta accggcaatt 240
catctcttg tccccatccgt ggggtggcta tatacagcaa agacaatagc ataagaattg 300
gttccaaagg agacgtttt gtcataagag agcccttat ttcatgttct cacttggaat 360
gcaggacctt ttttctgacc caaggtgcct tactgaatga caagcattca aatggactg 420
ttaaggacag aagcccttat agggccttaa tgagctgcc tgtcggtgaa gctccgtccc 480
cgtacaattc aagattgaa tcggtgtt ggtcagcaag tgcatgtcat gatggcatgg 540
gtctggctaac aatcgaaatt tcagggtccag ataatggagc agtggctgtt taaaataca 600
acggcataat aactgaaacc ataaaaagtt ggaggaagaa aatattgagg acacaagagt 660
ctgaatgtgc ctgtgtaaat ggttcatgtt ttactataat gactgatggc ccgagtgatg 720
ggctggcctc gtacaaaatt ttcaagatcg aaaaggggaa ggttactaaa tcaatagagt 780
tgaatgcacc taattctcac tatgaggaat gttcctgtta ccctgataacc ggcaaagtga 840
tgtgtgtgtc cagagacaac tggcatgtt cgaaccggcc atgggtgtct ttcatgtt 900
acctggatta tcaaataatgga tacatctgca gtggggttt cggtgacaac ccgcgtcccg 960
aagatggAAC aggtagctgt ggtccagttt atgttgatgg agcaaaacggg gtaaaggat 1020
tttcataatag gtatgtttaat ggtgtttggta taggaaggac caaaaagtcac agttccagac 1080
atgggtttga gatgattttgg gatcctaataat gatggacaga gactgatagt aagttctctg 1140
tttaggcaaga tgggtgtggca atgactgtt ggtcagggta tagcggaaat ttcgttcaac 1200
atccctgagct aacagggtca gactgtatga ggccgtgtt ctgggttgaa ttaatcagg 1260
gacgaccta aaaaaaaaaaca atctggacta gtgcgagcag cattttttt tggcgttgc 1320

-continued

atagtgtatac tgttagattgg tcttggccag acgggtctga gttgccattc agcattgaca 1380
agtagtctgt tcaaaaaact ccttgtttct act 1413

```
<210> SEQ ID NO 25
<211> LENGTH: 1413
<212> TYPE: DNA
<213> ORGANISM: unknown
<220> FEATURE:
<223> OTHER INFORMATION: synthetic
```

<400> SEQUENCE: 25

acgcaaaggc ggggttaaaa atgaatccaa atcagaaaaat aacaaccatt ggatacatct
gtctggtagt cggactaatt agcctaataat tgcaaataagg gaatataatc tcaatatgg
tttcgcattc aatccaaacc ggatcacaaa atcatacagg catatgcaat cagaataataa
ttacttataa aaatagtaca tgggtgaaag atactactag cgtgatacta accggcaatt
ctagtctatg tccgattagg gggtggtcta tatactctaa agacaatagt atacggatag
ggtctaaggg agacgtttc gtaattaggg aaccgtttat aagttgttca catctagagt
gttaggacctt ttttctgaca caaggcgac tatataaacga taagcattct aacggatcag
ttaaggatag gtcacccattt agggcactta tgtcatgtcc cgttagggaa gcccttagtc
cataacaatag tagattgaa tccggtgcattt ggtccgcgtat cgcatgtcac gacggaaatgg
ggtggttgac tataggattt acgggacccg ataacggagc cggtggcgta ctgaaatata
acggtataat taccgaaact attaagagtt ggctttaaaaa aatattgtcgt acacaagagt
ccgaatgcgc atgcgttaac ggatcatgtt ttacaattat gactgacgga cctagcgacg
ggtagcgtc atacaaaattt tttaaaatcg aaaaaggcaa ggttactaag tcaatcgagt
taaacgcacc taattgcatt tacgaagagt gttcatgttta tcccgataacc ggaaagggtt
tgtgcgtttt tagggataat tggcacggtt cgaacagacc ttgggtgtca ttcgatcaaa
atctagacta tcaaatacgat tatatatgtt gcccgtataat cctagaccag
aggacggtac aggacgttgtt ggaccgggtt acgttgcacgg cgctaacggc gttaagggtt
ttatgttagt atacggcaat ggcgtatggat tcggtaggtt taagtccat agttcttagac
acggatttga aatgatatagg gatcttaacg gatggaccga aaccgactcg aagtttagcg
ttaggcaaga cgtatgcgtt atgaccgatt ggtccgggtt tagcggatca ttcgtgcac
atccagagtt aaccggatgtt gattgtatgc gaccatgttt ttgggttgag ttgatttaggg
ggagacccgaa agagaaaaact atatggacta gcgcgagcag cattttttt tgcggcgtga
atagtgatac tgcgtatgg tcttggccag acgggtgcgtt gttgcattc agcattgaca
agtatgttgtt cttttttttt act 1413

<210> SEQ ID NO 26
<211> LENGTH: 1026
<212> TYPE: DNA
<213> ORGANISM: Influenza virus

<400> SEQUENCE: 26

agcgaaaagca ggttagatatt gaaagatgag tcttctaacc gaggtcgaaa cgtacgtact	60
ctcttatcatc ccgtcaggcc ccctcaaagc cgagatcgca cagagacttg aagatgtctt	120
tgcagggaag aacactgtat ttgaggttct catggaatgg ctaaagacaa gaccaatcct	180
gtcacctctg actaagggga ttttaggatt tgtgttcacg ctcaccgtgc ccagtgagcg	240
aggactgtcgac cgtagacgct ttgtccaaaa tgcccttaat ggaaacgggg atccaaataa	300

```

catggacaaa gcagttaaac tgcataaggaa gctcaagagg gagataacat tccatgggc      360
caaagaatac tcactcagtt attctgctgg tgcaacttgcg agttgtatgg gcctcatata      420
caacaggatg ggggctgtga ccactgaagt ggcatttgc ctggtatgtg caacctgtga      480
acagattgtc gactcccagc atcggtctca taggcaaatg gtgacaacaa ccaatccact      540
aatcagacat gagaacagaa tggttttagc cagcactaca gctaaggcta tggagcaaat      600
ggctggatcg agtgagcaag cagcagaggc catggaggtt gctagtccagg cttagacaaat      660
ggtgcaagcg atgagaaccca ttgggactca tcctagtcg agtgctggc tgaaaaatga      720
tcttcgttcaa aatttgcagg cctatcgaa acgaatgggg gtgcagatgc aacggttcaa      780
gtgatcctct cgctattgcc gcaaataatca ttgggatctt gcacttgaca ttgtggattc      840
ttgatcgtct tttttcaaa tgcatttacc gtcgctttaa atacggactg aaaggagggc      900
cttctacgga aggagtgcga aagtctatga gggagaataa tgcggaaaggaa cagcagagtg      960
ctgtggatgc tgacgttgtt cattttgtca gcatagagct ggagtaaaaa actaccttgt      1020
ttctac                                              1026

```

<210> SEQ ID NO 27

<211> LENGTH: 890

<212> TYPE: DNA

<213> ORGANISM: Influenza virus

<400> SEQUENCE: 27

```

agcaaaaacca gggtgacaaa gacataatgg atccaaacac tgcataaggc tttcaggtag      60
attgctttct ttggcatgtc cgcaaacgag ttgcagacca agaacttaggt gatgecccat      120
tccttgatcg gcttcggcga gatcggaaat ccctaaaggagg aaggggcagc accctcggtc      180
tggacatcga gacagccaca cgtgctggaa agcagatagt ggagcggatt ctgaaagaag      240
aatccgtatca ggcacttaaa atgaccatgg cctctgtacc tgcgtcgct tacctaactg      300
acatgactct tgaggaaatg tcaaggact ggtccatgct catacccaag cagaaagtgg      360
caggccctct ttgtatcaga atggaccagg cgatcatgg taagaacatc atactgaaag      420
cgaaacttcag tgcgtttttt gacccggctgg agactctaattt attgctaagg gcttcaccg      480
aaggaggagc aattttggc gaaatttcac cattgccttc tcttccagga catactgtg      540
aggatgtcaa aatgcagtt ggagtccctca tcggggact tgaatggaaataacacag      600
ttcgagtctc tgaaactcta cagagattcg ctggagaaag cagtaatgg aatggagac      660
ctccactcac tccaaaacag aaacggaaaa tggcgggaaac aatttaggtca gaagttgaa      720
gaaataagat gttgtattga agaagtggaa cacaaactgaa agataacaga gaatagttt      780
gagcaataaa catttatgca agccttacat ctattgttg aagtggagca agagataaga      840
actttctcg ttcagcttat ttaataataa aaaacaccct tgcgttctact                  890

```

<210> SEQ ID NO 28

<211> LENGTH: 890

<212> TYPE: DNA

<213> ORGANISM: unknown

<220> FEATURE:

<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 28

```

agcaaaaacca gggtgacaaa gacataatgg atccaaacac tgcataaggc tttcaggtag      60
attgctttct ttggcatgtc cgcaaacgag ttgcagacca agaacttaggt gatgecccat      120

```

US 9,476,032 B2

157

158

-continued

tccttgaccg actgagacgg gatcagaataat cccttagggg caggggatcg accctaggcc 180
tagacatcgaa acccgcaact agggccggaa agcagatctgt ggagcgtata ctgaaagagg 240
agtcccgacgaa agcgcttaag atgactatgg ccagcgtacc cgcttagtccg taccttaccc 300
atatgacact cgaagagatg tcacgcgatt ggtctatgtct aatccctaag cagaaagtgg 360
ccggacacctt atgtatacgg atggaccagg cgattatggaa caaaaacatt atccttaag 420
cgaacttttc cggtatattc gatcgcttag agactctgtat actgttgctgt gcattcacag 480
aagggggacg aatttgtggc gaaatttcac cattgccttc tttccagga catactgctg 540
aggatgtcaa aaatgcgtt ggagtcctca tcggggact tgaatggaat gataaacacag 600
ttcgagtcctc tgaaactcta cagagattcg ctggagaag cagtaatgag aatggggac 660
ctccactcac tccaaaacag aaacgagaaa tggcggaaac aatttaggtca gaagttgaa 720
gaaataagat gggtgattga agaagtggaa cacaactgaa agataacaga gaatgtttt 780
gagcaaataa catttatgca agccttacat ctattgttg aagtggagca agagataaga 840
actttctcgat ttca gtttat ttaataataaa aaaacacccct tggtttact 890

<210> SEQ ID NO 29
<211> LENGTH: 7152
<212> TYPE: DNA
<213> ORGANISM: Human rhinovirus 89

<400> SEQUENCE: 29

ttaaaactgg gagtggggtg ttcccactca ctccacccat ggggttgtt actctgttat
tacggtaact ttgtacgcga gttttccca ccctccccca taatgtact tagaagtgg 1200
tacaatatgaa ccaataggta acaatcatcc agactgtcaa aggtcaagca cttctgttcc 1800
cccggtcaat gaggatatgc tttacccaag gcaaaaacct tagagatcgt tatccccaca 2400
ctgcctacac agagcccagt accattttg atataattgg gttggcgct ccctgcaaac 3000
ccagcagtag acctggcaga tgaggctgga cattcccccac tggcgacagt ggtccagcc 3600
ggcgtggctgc ctgctcaccc ttcttgggtg agaagcctaa ttattgacaa ggtgtgaaga 4200
gccgcgtgtg ctcagtggtc ttccctccggc ccctgaatgt ggctaaccctt aaccctgcag 4800
ccgttgcacca taatccaaatg ggtttgcggt cgtaatgcgt aagtgcggga tgggaccaac 5400
tactttgggt gtccgtgtt cctgttttc ttttatttttgc attttatggt gacaatttt 6000
agtgtataga ttgtcatcat gggtgaccaa gtatctagac aaaatgttgg gacacactcc 6600
acacaaaattt cagtgagcaa tggatctagc ttaaatttttcaacatcaa ttattttaaa 7200
gacgcagctt caagtgggtc ttcttagattt gatTTTCTC aagacccttag taaattttact 7800
gatccgttta aagatgtgtt agaaaagggt attccaaacac ttcaatcacc aacagttgaa 8400
gcttgggtt attcagacag actaatacag ataaccggag gagattccac tataacatcc 9000
caagatactg ccaaattttttt ggtgtgtggc catcataactt gactccagat 9600
gatgcgactg ctattgacaa acccacacaa cctgatacat catccaaacag attctacacc 10200
ttggacagtc gttctggac attcgcctca tctggatggt ggtggaaattt gcctgtatgcc 10800
cttaaaaaca tgggtattttt tggtaaaaaat atgttttacc attttctagg gagatgttgg 11400
tacacaatac atgttacaatg taattctagc aagtttcatc agggtttattt aatagttgcc 12000
ggccatcccaag aacatcaattt ggcatctgca acaagttggaa atgtatcgtt cgggtacaat 12600
cacacccacc caggtgagca aggttagagaa gtagtaccat cacggacatc tagtgataat 13200
aaaagaccta gtgtatcgtt gttgttaaat tttgtatggaa cattacttgg taacttacct 13800

US 9,476,032 B2

159**160**

-continued

atttatcccc accaatacat taatctaagg actaacaatt cagctaccct tattttacct 1440
 tatgtcaatg ctgtaccaat ggactctatg ctttagacata ataattggag cttggttata 1500
 atcccaatat gccccttca ggtccaacct ggggggacac aatccatacc tataacagta 1560
 tcaattagcc ctatgtttc agaattttca gggccaagaa gtaaggttgt gttagtacc 1620
 actcagggtt taccagttat gttaacacct ggatctgggc aattcttaac aactgtatgt 1680
 actcaatccc catcagcggtt tccatacttc cacccgacca aggaatattatacctgga 1740
 caagtttagga atttaattga aatgtgccaa gttgacacac tcattctgt taacaataca 1800
 caggaaaatg taagatctgt gaatatgtac actgttgatt tacgcacaca agttgattta 1860
 getaaagaag tctttctat accagtagat attgcctcac aacctttagc cactactctc 1920
 ataggagaac ttgcaagcta ttacacacac tggactggta gtctgcgcct tagctttatg 1980
 ttttgggtt ctgctagctc tactttgaaa ctattaattg catacactcc tcctgggtt 2040
 gggaaaaccta aatccaggag agaagccatg cttggcacac atttagtgtg ggatgtgggg 2100
 ttgcagtccca ccgcctcact agttgtacca tgggttagtg ctagccattt tagattcact 2160
 acacctgaca catattccctc agctggttat attacatgtc ggtaccagac caactttgt 2220
 gtacctgata gtactccaga taacgccaaa atgggtgca tggtttgc atgcaaagat 2280
 ttttgcattaa gattagccag agatactaac ctacacacac aagaaggagt actcacacaa 2340
 aacccagttt aaaatttat agatagtgtt ttaaatgaag ttcttgcgtt gccaaatatc 2400
 caacctgca catctgtgtc aagtcatgca ggcgcgtcat tggatgctgc gggaaaccgga 2460
 cacaccagct ctgttcaacc tgaagatatg attgaaacta gatatgttat aactgtatca 2520
 acaagggatg aaacaagttt tgagagttt ttaggttagt cagggtgtat cgctatgata 2580
 gaatttaata caagtagtga taaaactgaa catgataaaa ttggtaaagg attcaaaca 2640
 tggaaaggta gtcttcaaga aatggcacaa atcagaagaa aatatgaatt attcacat 2700
 acaagatttt attcagagat aacaatagtc actgcagccg cagctcaagg aaatgtatgt 2760
 ggacatatacg tatttcaatt tatgtatgtt ccccccaggag cacctgtccc cgaaaaacgt 2820
 gatgattaca catggcaatc aggaacaaat gcatctgtgt tctggcaaga aggacaacca 2880
 taccggatgt tcacaatccc ttttatgac attgcattcg cctattacat gttttatgt 2940
 ggtttatgtt gtgatagtgc agcatcaaaa tacgggtctg tagtcaactaa tgatatggg 3000
 accatatgtt ttagaatatg gacatccaac caaaaacacg attcaaataat tggatgcgc 3060
 atttaccaca aggccaaaca tataaaagca tgggtgcctc gcccaccaag ggctgttgcc 3120
 tatcaacaca cacactcaac caattacata ccatccaatg gtgaggccac aactcagatt 3180
 aaaaccagac ctgtatgttt taccgttaca aacgtcgac catctgtat gttttatgt 3240
 gtggtaact taatctatag aaatcttcat ctctttaattt ctgtatgttga tgattctatt 3300
 cttgtatcat actccagtgta tctaatcata tatcgaacaa acactgaagg taatgtatgt 3360
 atccctaattt gtgattgcac tgaatgtaca tattactgcc accacaaaga taggtat 3420
 cctatcagag ttactgcaca tgattggat gagattcaag aatcagaata ttacccaaa 3480
 catatccaaat ataatctctt gattggagag ggtccttgc aaccaggaga ttgtggagga 3540
 aaactattgt gtaaaacatgg tggatgtatgtt atgattacag ctggagggtga aggtcacgtt 3600
 gcttttatttgc acctgagaaa attccagtgatgt gctgaggagc aagggttac tgattatgt 3660
 gaacatcttgc tcaagtctt tgggttaggc ttcgttagaca gcatcaaaca acaggtaaac 3720

-continued

-continued

atagataaaa agaccaaaga catcacaaaa cttaggaaag caattgtatg atatggatt 6180
 gatggctta tggttacttt tctgaaagat gaacttagaa agaaggaaaa aataaaagat 6240
 ggaaagacta gagtcataga agctaatagt gtgaatgata ctgtgttatt cagaagtgt 6300
 ttggaaatc ttttctctgc tttccacaaa aaccaggta tagtcaactgg ttcagcagta 6360
 gggtgtgacc ctgaagtatt ctggtcaact atacctctca tgcttagatgg agaatgtta 6420
 atggctttt attattcaaa ctatgatggt agcctacatc ccgtttgggt taaatgtctt 6480
 agtatgtct tagaagacat aggttctcc tctcaactta ttaaccagat ctgtactct 6540
 aaacatatat acaaatactaa gtattatgaa gtggaaaggag gtatgccatc tggatgtgct 6600
 ggtactagta ttttaatac aataatcaac aatattatca ttagaacttt ggtactagat 6660
 gettataaga acatagatct agataaactg aaaatcttag catatggga tgatgtcatc 6720
 ttttcttata attttaact tgacatggca gttcttgcca aagaaggaga aaaatatgga 6780
 ctaacaatca cccctgtga taagtctgtat gtttccaag aattgaccta taaaaatgt 6840
 acttttctta aaagaggatt cagagctgtat gagcgccact ctttccttat acaccctacc 6900
 tttcctgtgg otgagattca tgactccatc agatggacca aaaacccttc atgtatgcag 6960
 gaacacgtgc tatctttgtg tcatttaatg tggcataatg gtagacatgc ataccaggaa 7020
 ttcattaaag gtatacgcag tgtatctgcc ggtcgggcac tgtatatacc agcttatgaa 7080
 gttcttgaac atgaatggta tgaaaaattt tagatataaa actgttaaat atagtagtt 7140
 tattagttt at 7152

<210> SEQ ID NO 30
 <211> LENGTH: 7152
 <212> TYPE: DNA
 <213> ORGANISM: unknown
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 30

ttaaaactgg gagtgggttg ttcccactca ctccacccat gcggtgttgt actctgttat 60
 tacggtaact ttgtacgcca gttttccca ccctccccca taatgtact tagaagttt 120
 tacaatatga ccaataggta acaatcatcc agactgtcaa aggtcaagca cttctgtttc 180
 cccggtaat gaggatatgc tttaccaag gcaaaaacct tagagatcgt tatcccaca 240
 ctgcctacac agagccagt accatttttgc atataattgg gttggcgtct ccctgcaaac 300
 ccagcagtag acctggcaga tgaggctgga cattccccac tggcgacagt ggtccagcct 360
 gctgtggctgc ctgctcaccc ttcttgggtg agaagcctaa ttattgacaa ggtgtgaaga 420
 gccgcgtgtc tcagctgtgc ttctccggc ccctgaatgt ggctaaacctt aaccctgcag 480
 cgcttgccca taatccatg gtttgcgggt cgtaatgcgt aagtgcggga tgggaccaac 540
 tactttgggt gtccgtgttt cctgttttc ttttgcgttgc attttatggt gacaatttat 600
 agtgtataga ttgtcatcat gggagcacaa gtgtctagac aaaacgtagg tacacatagt 660
 acacaaaact cagtgtctaa cggatctagt cttattttttaatattaa ttatattaaa 720
 gacgctgcacat ctagcggagc tagtagactt gattttagtc aggatccatc taaatttacc 780
 gatccagttt aagacgtact cgaaaaaggt atacctacat tgcaatcacc tacagtcgaa 840
 gcatgcggat actctgtatag attgatacaa attacttaggg gggatagtcgatattactgt 900
 caggataccg ctaacgcgt agtgcgcatac ggagttggc catcttatct tacacctgtat 960

-continued

gacgcaccc caatcgataa acctcacaca cctgatacta gttctaatag attctataca 1020
cttgattcta gatcatggac tagtgctagt tcagggttgt ggtggaaatt gcctgacgca 1080
cttaaaaata tgggttattt cggtgagaat atgttttacat atttcttagg taggtcagg 1140
tatactatac acgttcaatg taatagtct aaatttcacc aaggcttatt aatcggtcc 1200
gctatacccg aacaccattt agctagcgct acatcaggta acgttagcggt agggtataat 1260
catacacatc caggcgaaca gggtagggaa gtcgtaccat cttagaacatc tagcgataat 1320
aaacgaccat ctgacgatag ttggcttaat ttcgatggta cactattagg taatctacct 1380
atataccac atcaatatat taatctttaga actaataatt ccgcttacact tatactgcca 1440
tacgttaacg cagtgcctat ggatagtatg cttagacata ataattggtc attagtgatt 1500
ataccgatat gtccattgca agtgcacacct ggaggtacac aatctatacc tattaccgtt 1560
agtatttagtc ctatgtttag cgaattttagc ggaccttagat ctaaggtagt gtttagtact 1620
acacagggtt accccgttat gcttacacct ggatcagggtc aatttttgc tactgacgat 1680
acacaatcac ctatgtcattt cccatattt caccctacta aagagatttt tatacccggt 1740
cagggttagga atctaatcga aatgtgtcaa gtcgatacac ttataccgt taataataca 1800
caggaaaacg ttagatccgt taatatgtat accgttgatt tgctacaca agtcgatcta 1860
gctaaagagg tttttcgat accagtcgat atcgcttagt aaccatttagc tactacactt 1920
ataggcgaac tcgcatcata ttatacacat tggaccggat cacttagatt ttctttatg 1980
ttttgcggat ccgtctatg tacacttaaa ctgttaatcg catatacacc tccaggagtg 2040
ggtaagecota aatctagacg cgaagctatg ttaggtacac atctagtttggacgttaggg 2100
ttacaatcta ccgtctatg agtcgtacca tgggttagcg catcacattt tagatttact 2160
acacctgata cttattctag tgccggatata aacatgtt ggtatcaaac taatttcgta 2220
gtgccagata gtacacctga taacgctaaa atggttgtt gtttagtgcg atgtaaaat 2280
ttttgtctta gattagctag ggataactat ttgcatacac aagagggagt gtttacacaa 2340
aatccagtcg aaaattatcgat ctagatcgta cttaaacgaa tgtttagtgcg acctaataata 2400
caacctagta ctagtgcgtc atctcagctgc gcaccagcat tagacgcagc cgaaaccgg 2460
catacatcta gtgtgcacc tgaggatatg atagagacta gatacgttt taccgatcag 2520
actagagacg aaacatcaat cgaatcattt ttaggttagt cagggtgtat cgctatgata 2580
gagtttataa ctagttccga taaaaccgaa cacgataaga taggtaaagggtt gttttaaact 2640
tggaaagtgtt cattgcagga aatggcacaa attagacgta aatacgaatt gtttacatata 2700
actagattcg atagttagat tactatcgat accgcagccg cagcacaagg taacgattca 2760
ggtcataatg tgttacaatt tatgtacgta ccaccagggtg caccagtacc cgaaaaacgc 2820
gtgattata catggcaatc cggtaactac gctagcgat tttggcaaga gggacaacca 2880
tacccatgtat ttactatacc ttttatgtca atcgcttagt catattatata gttttacgac 2940
ggatacgtat ggcattctgc cgcataaaa tacggatctg tagtactaa cgatatgggt 3000
actatatacgat tttagatcgat tacatctaat caaaaacacg attctaatat cgtttgcgt 3060
atatactata aagctaaaca tattaaggca tgggtgcataa gaccacccatg agccgttgc 3120
tatcaacata cacatagtagtac taattatata ccatctaaacg gtgaggctac tacacaatt 3180
aaaactagac ctgacgtttt taccgttact aacgttaggtc catctagttt gttcgatcac 3240
gttaggtatc tgatataatg aaatttgcattt ctattcaattt ccgatatttgcgatcgat 3300
ctcggtatg attcttagcgat tctgattata tatagaacta ataccgaagg taacgatgtg 3360

-continued

atacctaatt	gcgattgtac	cgaatgtact	tattattgtc	atcataaaga	taggtatTTT	3420
ccgatacgCG	ttaccgcaca	cgattggTAC	gaaatacagg	aatcttagat	ttatCCTAAA	3480
catatacaat	ataatctgtt	aataggcGAA	ggtccatgcG	aaccaggcGA	ttgcggagGT	3540
aagttattgt	gtaaacacGG	agtgtatAGGT	atgattacCG	caggGGGGGAGA	gggacacGTT	3600
gcgtttatcg	atttgcgaaa	atttcaatgc	gcagagGAAC	aggggttATC	cgattacGTT	3660
gagcatttag	ggcaagtgtt	cggagtCGGA	ttcgttgatt	caattaaACA	acaggTTAAT	3720
tttattaatc	ctacatctaa	aatcgatCT	aaagtgtataa	aatggTTACT	tagaatcgTT	3780
agcgctatga	taattatGgt	taggaattCT	agcgatCCAC	aaaccgtAAT	cgctcacACTT	3840
acactattAG	ggtgTTcagg	ttcacCTTGG	cgatttCTTA	aagagaaATT	atgcgcATGG	3900
ttgcaattgt	oatacgttca	taaacaatCC	gatagttGGC	ttaaaaAAATT	taccgaAGCA	3960
tgtaacgccc	ctaggggact	cgaatggata	gggcaaaaaAA	tatctaaATT	tatcgatTTG	4020
attnaatctA	tgttaccaca	agcgcattG	aaaatcgatt	atcttactAA	gcttaaacAA	4080
ttgaacttac	tcgaaaaaca	aatcgaaACT	attagactCG	caccagCTAG	tgtgcaAGAG	4140
aaaattttta	tagagattaa	tacattacac	gatctatCAC	ttaaattCTT	accattatac	4200
gttagcgaag	ctagacggat	taaaaaatcta	tatattaAGT	gttctaACGT	aattaaggGA	4260
ggtaagcgtA	acgaacccGT	tgccgtactt	atacacGGAT	caccaggTCAC	cggtaaAGTC	4320
ttagcgcactA	gcgtactcgc	tagaatgCTT	acagtcGAA	ccgatataATA	ctctttacCA	4380
cctgatccta	aatattcga	cggatacgt	caacaatCCG	tagtgattat	ggacgatatt	4440
atgcAAATC	ctagtggta	ggatATGACA	ttgtttgtc	aatggTATC	tagcgtacCA	4500
tttataccac	ctatggccGA	tttaccCGAT	aagggtAAACG	cttttacATC	taaattcGTA	4560
ctcgcatcaa	ctaattcatac	attgcttaca	ccacctaccG	ttagttcaCT	accagctatG	4620
gttagacggT	tttatttGTA	tctagatata	caggTTAAA	aagagtatCT	gttagacGGT	4680
aagttagata	tcgcttaatC	ttttagacca	tgcgatGTTA	atattaaAT	cggtAAcGCT	4740
aaatgttGTC	catttatATG	cggtaaggCA	gtcgaatttA	aagataggAA	ttcatgtACT	4800
acactatcac	tatcacaatt	gtattcacat	attaaggAAG	aggatagacG	tagatCTAGT	4860
gccccacaAG	ctatggAAc	tatTTTCAG	ggtatcgatC	tgcaatCACC	accaccacCA	4920
gtatAGCCG	atctacttag	atccgttAA	acacCTGAGA	taattaAGTA	ttgtcaggat	4980
aataattGGA	tcgttaccAGC	cgaatgttca	atcgaacGCG	ttatAGGGAT	agcgaatATG	5040
acaatcggta	taatcgctaa	cgttgtgtca	atcgtaggCG	ttatatatat	tatataAA	5100
ttgtttgtA	cattgcaggG	accatactCT	ggcgaaccta	aacctaaATC	taggcacCT	5160
gaaagacgCG	tagtgacaca	gggacctGAA	gaggaattCG	gttaggtatt	gtcttaAACAT	5220
aattgttGCG	tagtgactac	cgataaggGT	aagtTTACCG	gatttaggtat	atacgatcAG	5280
gttatgggt	tacctacaca	ttccgatcca	ggatctgaga	tactcgtaga	cgaggTTAG	5340
gttaaggTTA	gcgattcata	cgatctacat	aatcacGAAG	gcgttaAGTT	agagattacc	5400
gtatgttAAGC	ttatacgtAA	cggAAATTT	aaagatatac	gaaaatATCT	accatcacGC	5460
gaagacgatt	atcctgcata	taatctcgca	ctattagcga	atcaagacGA	acctacAA	5520
attagcgtAG	gcgatGCCGT	atcatacGGT	aatataCTGT	tatccggTAC	taataccGCT	5580
agaatgatta	agtatcatta	tcctactaaa	gccggatatt	gcccggaggGT	gttatataAA	5640
gtcggatcta	tactcggtat	acacgttagc	ggtaacggta	gggacggatt	ttccgctatG	5700

US 9,476,032 B2

169

170

-continued

ttaacttaat cttatttcgg tgagacacag ggattgatta ctaaagagggtt acccgatcc	5760
gttaaaaatc taccttagcgt acacggttgt tctaaaacta gattgcaacc tagtgtgttt	5820
cacgaegttt ttccaggtac taaggaacct gccgtactat cttctaacga tcctagactc	5880
gaaaccgatt tcgatagtgc attattctct aaataaagg gtaatcctgc atgtcagggtt	5940
acaccacata tgaaaatcgc agtgcgcacat tacgctgcac aattgtctac acttgatatt	6000
aatccacaac cactatcact tgaggaatcc gtattcggta tcgaagggtt agaggcactt	6060
gatcttaata ctagtgcgg attccacatac gttagtcgtt gtattaaaaa aaaagatcta	6120
atcgataaaa aaactaagga tattactaaa ttgcgaaaag cgatagacga atacggatc	6180
gatctaccta tggttacatt ccttaaagac gaacttagaa aaaaagagaa aattaaagac	6240
ggtaagacta gggttatcga agctaactct gttaacgata ccgtacttt tagatccgta	6300
ttcggtaatc tattttccgc attccataaa aatccaggta tcgttaccgg atccgcagtc	6360
ggatgcgatc cagaagtgtt ttggtctact ataccactta tgtagacgg tgagtgtctt	6420
atggcattcg attattctaa ttacgatggg tcattgcac ccgtatgggt taaatgtcta	6480
tctatgttac tcgaagatata agggtttagt tcacaattga ttaatcagat ttgtactct	6540
aaacatataat ataaatctaa atattacaa gtcgaagggg gtagccatc aggttgcgca	6600
ggtacatcta ttttaatac tataattaat aatattataa ttagaacatt agtgttagac	6660
gcatataaaa atatcgatct agataagctt aagataactcg catacgaga tgacgttata	6720
ttctcatata attttaact tgatatggcc gtactcgcta aagaggaga aaaatacgg	6780
cttacaatta cacctgccga taagtctgac gttttcagg aattgactta taaaacggtt	6840
acattcccta aaaggggggtt tcgcgcagac gaacgacatt cttttcttat acatcctaca	6900
tttcccgttgc cgaaatacaca cgattcaatt agatggacta aaaatcctag ttgtatgcaa	6960
gaacacgtac ttagtctatg tcatctttag tggcataacg gtagacacgc ttatcaggaa	7020
tttattaagg gtattagatc cgtagtgcgc ggttagggcac tatataacc cgcatacgaa	7080
gtactcgaac acgaatggta cgaaaaattt tagatataaa actgttaaat atagctagtt	7140
tattagttt at	7152

<210> SEQ ID NO 31
 <211> LENGTH: 7212
 <212> TYPE: DNA
 <213> ORGANISM: Human rhinovirus 14

<400> SEQUENCE: 31

ttaaaaacagc ggatgggtat cccaccattc gacccattgg gtgttagtact ctggtactat	60
gtacctttgt acgcctgttt ctcccccaacc acccttcctt aaaattccca cccatgaaac	120
gttagaagct tgacattaaa gtacaatagg tggcgccata tccaatgggtg tctatgtaca	180
agcaacttctg tttcccgagga gcgaggtata ggctgtaccc actgccaaaa gcctttaacc	240
gttataccgc aaccaactac gtaacagtta gtaccatctt gttcttgact ggacgttgc	300
tcagggtggat tttccctcca ctatgtttgtt cgatgaggctt aggaattccc cacgggtgac	360
cgtgtccctag cctgcgtggc ggccaaacca gcttatgtcg ggacgcccctt ttaaggacat	420
ggtgtgaaga ctcgcattgtg cttgggttg agtccctccgg cccctgaatg cggctaacct	480
taaccctaga gccttatgcc acgatccagt ggttgcgttgc tctcatatgt tcttatggc	540
acgggaccga ctactttggg tgcgtgtttt cttcatatgt tcttatggc	600
acagcatata tatacatata ctgtgatcat gggcgctcag gtttctacac agaaaagtgg	660

-continued

atctcacgaa aatcaaaaaca ttttgaccaa tggatcaa at cagacttca cagttataaa	720
ttactataag gatgcagcaa gtacatcatc agctggtaa tcactgtcaa tggaccatc	780
taagttaca gaaccaggta aagatctcat gcttaagggt gcaccagcat tgaattcacc	840
caatgttgag gcctgtggtt atagtgtatc agtacaacaa atcacactcg ggaattcaac	900
aataacaaca caagaagcag ccaacgctgt tgtgtgttat gctgaatggc cagagtacct	960
tccagatgtg gacgctagtg atgtcaataa aacttcaaaa ccagacactt ctgtctgttag	1020
gttttacaca ttggatagta agacatggac aacagggtct aaaggctggt gctggaaatt	1080
accagatgca ctcaaggata tgggtgtgtt cgggcaaaac atgttttcc actcaactagg	1140
aagatcaggta tacacagtt acgttcagtg caatgccaca aaattccata gcggttgtct	1200
actttagtt gtaataccag aacaccaact ggcttcacat gagggtggca atgtttcagt	1260
taaatacaca ttcacgcac caggtgaacg tggatagat ttatcatctg caaatgaagt	1320
ggggggccct gtcaaggatg tcataataaa tatgaatggt acttttattag gaaatctgct	1380
cattttccct caccagttca ttaatctaag aaccaataat acagccacaa tagtgatacc	1440
atacataaaac tcagtagcca ttgattcaat gacacgtc acacatgtct cactgtatgg	1500
catccctatt gccccttta cagtagccac tggagcaact coctcaactcc ctataacagt	1560
cacaatgca cctatgtgca ctgagttctc tggataagg tccaaatgca ttgtgccaca	1620
aggtttgcac actacaactt tgccgggtc aggacaattt ttgaccacag atgacaggca	1680
atcccccaact gcaactgcca attatgagcc aactccaaga atacacatc tagggaaagt	1740
tcataacttgc ttagaaatta tacaggtaga tacactcatt cctatgaaca acacgcatac	1800
aaaagatgag gttaacagtt acctcatacc actaaatgca aacaggcaaa atgagcaggta	1860
ttttggaca aacctgttta ttggatgg ggtcttcaaa actactcttc tgggtgaaat	1920
tgttcagtagc tatacaccatt ggtctggatc acttagattc tcttcgatgt atactggcc	1980
tgccttgc cgtctaaac tcactctacg atacaccccg cctggcgtc gtggccaca	2040
ggacaggaga gaagcaatgc taggtactca tggatgtgg gatattggc tgcaatccac	2100
catagtaatg acaataccat ggacatcagg ggtgcagttt agatatactg atccagatac	2160
atacaccagt gctggtttc tatcatgttg gtatcaaact tctttatc ttccccaga	2220
aacgaccggc caggctact tattatcatt cataagtgc tgcactgatc ttaagcttag	2280
gctgatgaaa gatactcaaa ctatctaca gactgtgc ctcactgaag gcttaggtga	2340
tgaattagaa gaagtcatcg ttgagaaaac gaaacagacg gtggcctcaa tctcatctgg	2400
tccaaaacac acacaaaaag tccccatact aactgcaac gaaacagggg ccacaatgcc	2460
tgttcttcca tcagacagca tagaaaccag aactacccat atgcacttta atgggtcaga	2520
aactgatgta gaatgtttt tgggtcgatc agcttgcgtg catgtactg aaatacaaaa	2580
caaagatgct actggaatag ataatcacag agaagcaaaa ttgttcaatg attggaaaat	2640
caacctgtcc agccttgtcc aacttagaaa gaaactggaa ctcttcactt atgttagtt	2700
tgattctgag tataccatac tggccactgc atctcaaccc tatttcactt atgttcaag	2760
caatttgggtg gtccaaagcca tggatgttcc acatggtgcc cggaaatcca aaagagtggg	2820
cgattacaca tggcaaaatgt cttcaaaccc cagtgtattt ttcagggtgg gggatacatc	2880
aaggtttagt gtgcctttagt taggattggc atcagcatat aattgtttt atgatggta	2940
ctcacatgt gatgcagaaa ctcagtagtg cataactgtt ctaaccata tgggttagtat	3000

-continued

ggcattcaga atagtaaatg aacatgatga acacaaaact cttgtcaaga tcagagttt	3060
tcacaggcga aagctcggtt aagcatggat tccaagagca cccagagcac taccctacac	3120
atcaataggg cgccacaatt atcctaagaa tacagaacca gtaattaaga agaggaaagg	3180
tgacattaaa tcctatggtt taggacctag gtacgggtgg atttatacat caaatgttaa	3240
aataatgaat taccacttga tgacaccaga agaccaccat aatctgatag caccctatcc	3300
aaatagagat ttagcaatag tctcaacagg aggacatggt gcagaaacaa taccacactg	3360
taaccgtaca tcaggtgttt actattccac atattacaga aagtattacc ccataatttgc	3420
cggaaagccc accaacatct ggattgaagg aagcccttat taccgaagta gatttcaagc	3480
aggagtgtatc aaaggggttg ggccggcaga gctaggagac tgcgggtggaa tttttagatc	3540
catacatggt cccattggat tgttaacagc tgaaggtagt ggatatgtt gtttgcgtt	3600
catacgtacag ttggagtgtt tgcgcagagga acaggggctg agtgattaca tcacagggtt	3660
gggttagagct ttgggtgtcg ggttcaactga ccaaacttca acaaaggctca cagaactaca	3720
agaagtgccg aaagatttcc tcaccacaaa agttttgtcc aaagtggtca aatggtttcc	3780
agcttttagtgc atcatttgc gaaatcatga tgacttggc actgttacgg ccactctagc	3840
actacttggc tgtgtatggat ctccctggag atttctgaag atgtacattt ccaaacttca	3900
tcaagggtcct tacattgaaa gacaagccaa tggatggatgg ttcaagaaagt ttaatgtatc	3960
atgtatgtc gcaaaagggtt gggatggat tgctataaaatc atttccaaac tgattgtatc	4020
gataaaaaac aaagtacttc cccaaagccaa agaaaaacta gaattttgtt gtaaactcaa	4080
acaacttgc atactagaga gacaataac caccatgtatc atctcgatc caacacagga	4140
aaaacgagag cagttgttca ataacgtatt gtgggtggaa caaatgtcgc aaaagtttgc	4200
ccccatttat gccgttgaat caaaaagaat cagggactc aagaacaaa tggtaaatttgc	4260
tatgcaattt aaaagtaaac aaagaactga accagtgtgt gtattaaatcc atggtacacc	4320
cggttctggt aaatcattaa caacatccat tggggacgt gcaattgcag aacacttcaa	4380
ttcagcagta tatttacttc caccagatcc caagcacttt gatggttatc agcaacagga	4440
agttgtgatt atggatgtatc tgaacccaaa tccagatggaa caggatataa gcatgttttgc	4500
tccaaatgggt tcttcagttt atttcttgc tccaaatggct agtttagata acaagggcat	4560
gttatttcacc agtaatttttgc ttcttagcctc cacaatttctt aacacactaa gcccccaac	4620
aatcttgaat cctgaagctt tagtcaggag atttgggtttt gacctagata tatgtttgc	4680
tactacccatc acaaagaatg gaaaactcaa tgcaggcatc tcaaccaaga catgcaaaaga	4740
ttggccatcaa ccatctaatt tcaagaaatg ttggccctca gtctgtggaa aagcttattag	4800
cttggtagac agaactacca acgttagtgc tagtgtgttgc caactggcgc cggcttattat	4860
aagtgtatcc aagagcaaaa tgcaaaattac agattccctca gaaacactgtt ttcaaggacc	4920
agtgtataaa gatttagaga ttgatgttttgc caacacacca ctttcagaat gtatcaacgc	4980
tttactgaaa tctgttagatt cagaagagat tagggatata tggatggatc agaaatggat	5040
tatacctgaa attccttacca acatagaaag ggctatgtatc caagccagca tgattattaa	5100
tactattctg atgtttgtca gtacatttgc tattgttttgc tgcattttata aattgtttgc	5160
tcaaaactcaa ggaccatatt ctggtaaccc gcctcacaat aaactaaaag ccccaacttt	5220
acggcccgatgttgcgttcaag gaccaacac agaatttgcgtt ctatccctgt taaggaaaaaa	5280
cataatgact ataacaaccc caaaggaga gttcacagggtt taggcatac atgatcgatc	5340
ctgtgtgata cccacacacg cacagcctgg tggatgtatc tggatgtatc gtcagaaaaat	5400

-continued

tagagttaa gataagtaca aatttagtaga tccagagaac attaatctag agcttacagt 5460
 gttgacttta gatagaaaatg aaaaattcag agatatcagg ggattttat cagaagatct 5520
 agaagggtgtg gatgccactt tggtagtaca ttcaaataac tttaccaaca ctatcttgc 5580
 agttggccct gtaacaatgg caggacttat taattttagt agcacccccca ctaacagaat 5640
 gattcgttat gattatgcaa caaaaactgg gcagtgtgga ggtgtgtgt gtgctactgg 5700
 taagatcttt ggtattcatg ttggcggtaa tggaaagacaa ggatttcag ctcaactaa 5760
 aaaaacaatat tttgttagaga aacaaggcca agtaatagct agacataagg ttagggagtt 5820
 taacataaat ccagtcaaca cggcaactaa gtcaaaaatta catcccagtg tattttatga 5880
 tgaaaaatcca ggtgacaagg aacctgctgt attgagtgac aatgatccc gactggaagt 5940
 taaaattgact gaatcattat tctctaagta caagggaaat gtaaatacgg aaccactga 6000
 aaatatgctt gtggctgttag accattatgc agggcaacta ttatcactag atatccccac 6060
 ttctgaacctt acactaaaag aagcattata tggagtagat ggactagaac ctatagat 6120
 tacaaccagt gcaggatttc cctatgttag tcttggatc aaaaagagag acattctgaa 6180
 taaagagacc caggacacag aaaagatgaa gttttatcta gacaagtatg gcattgactt 6240
 gcctctagtt acatataat aggatgaatt aagaagtgtt gacaaagtcc gattaggaa 6300
 aagtagatta attgaagcct ccagtttggaa tgattctgtt aacatgagaa tgaaactagg 6360
 caacctttac aaagcattcc atcaaaatcc cggtgttctg actggatcag cagtgggtt 6420
 tgatcctgat gtgttttgtt ctgtcatccc ttgcattatg gatggccacc tgatggcatt 6480
 tgattactct aattttgatg cctctttgtc accagtttgg tttgtctgtc tagagaaggt 6540
 ttgaccaag ttaggctttc caggctcttc attaattcaa tcaatttgc atacccatca 6600
 tatctttagg gatgaaatat atgtgggtga aggtggcatg ccctcagggt gttcaggaac 6660
 cagcatattc aattccatga tcaacaacat aatcattagg actttgatat tagatgcata 6720
 taaaggaata gatttagaca aacttaaaat cttagctac ggtgatgatt tgattgttc 6780
 ttatccctttaa gaactggatc cacaagtgtt ggcaactctt ggtaaaaatt atggactaac 6840
 catcacaccc ccagacaaaat ctgaaacttt tacaaaaatg acatggaaa acttgacatt 6900
 tttaaagaga tacttcaagc ctgatcaaca atttccctt ttgggttacc cagttatgcc 6960
 catgaaagat atacatgagt caatcagatg gacaaaggat cctaaaaaca cacaggatca 7020
 cgtcccgatca ttatgcattt tagcatggca ctcaggagaa aaagagtaca atgaattcat 7080
 tcagaagatc agaactactg acattggaaa atgtctaatt ctcccagaat acagcgtact 7140
 taggaggcgc tgggttggacc ttcttttagt taacaatata gacacttaat ttgagttagaa 7200
 gtaggagttt at 7212

<210> SEQ_ID NO 32
 <211> LENGTH: 7212
 <212> TYPE: DNA
 <213> ORGANISM: unknown
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 32

ttaaaaacagc ggtatgggtat cccaccattc gacccattgg gtgttagtact ctggtaactat	60
gtacctttgtt acgcctgtttt ctcccccaacc acccttcctt aaaattccca cccatgaaac	120
gttagaaagct tgacattaaa gtacaatagg tggcgccata tccaaatggtg tctatgtaca	180

-continued

agcacttctg tttcccagga gcgaggtata ggctgtaccc actgcacaaa gccttaacc	240
gttatccgcc aaccaactac gtaacagtta gtaccatctt gttcttgact ggacggttcga	300
tcaagggtggat ttccctcca ctagtttgtt cgatgaggct aggaattccc cacgggtgac	360
cgtgtcttag cctgegtggc ggccaaccca gcttatgctg ggacgcccctt ttaaggacat	420
ggtgtgaaga ctgcgtatgtt cttgggttgtg agtcctccgg cccctgaatg cggtcaacct	480
taaccctaga gccttatgcc acgatccagt gggtgtttaagg tcgtaatgag caattccgg	540
acggggaccga ctactttggg tgtccgtgtt ttcattttt cttcatatgt tcttatggc	600
acagcatata tatacatata ctgtgtatcat gggcgacaaa gtgtcaacgc aaaaatccgg	660
atcacacgaa aacccaaaaca tactgactaa cgatctaata cagactttt cagtgatcaa	720
ttattataaa gacgcccgtta gtacttagtc agccggtaaaa tcactatcta tggatccatc	780
taaatttacc gaacccgtta aggatcttat gcttaaaggc gcaccagcac ttaactcacc	840
taacgttgag gcatgccccat actctgtatag agtgcaacaaa attacattgg gtaatagtag	900
aattactaca caggaagccg ctaacgcgt cggttgttac gccgaatggc ccgaataactt	960
acccgatgtc gacgcgtatcg acgttataaa gacatctaaa cccgataacta gcgtatgttag	1020
attctataaca ctcgatagta agacttggac taccggtagt aagggtatgg gttggaaatt	1080
gcctgacgca cttaaggata tggggcgtatt cggtcagaat atgttttttc actcattttgg	1140
taggtcaggg tatacgtatc acgttcaatg taacgctact aaatttcaact cagggtgtct	1200
gttagtcgtt gtgatacccg aacaccaatt ggcatacacac gaagggggta acgttagcgt	1260
taagtataaca tttacacatc caggcgaaag aggtatcgat ctatcatccg ctaacgaagt	1320
ggggggaccc gttaaagacg ttatatacaa tatgaacggt acactgttag gtaatctgtt	1380
aattttcca catcaatcta ttaatctgtatc aactaataat accgctacaa tcgttataacc	1440
atatattaac tcagtgccaa tcgatagtagt gactagacat aataacgtta gtcttatgg	1500
gataccaatc gcaccattga cagtgcacac cggtgcaaca cctagtctac caattacgg	1560
tacaatcgca octatgttgc ctgagtttc cggttattagg tctaaatcaa tcgttaccaca	1620
aggggttacct actactacat tgcccgatc cggtcaattt ttgactactg acgataggca	1680
atcaccttagc gcattgccta attacgaacc tacaccacgt atacatatac tcggtaaggt	1740
gcataatctg ttagagatta tacaggtcgta tacactgata cctatgata atacacatac	1800
aaaagacgaa gtgaattcgt atctgtatcc gttaaacgct aataggcaaa acgaacaggt	1860
atccggtaacg aatctgttta taggcgtatgg cggttacaaa actacactgt taggcgaat	1920
cgttcaatac tatacataat ggtcagggttc acttagatcc tcatctatgt atacagggtcc	1980
agcactatct agcgctaaat tgacactcgc atatacacct cccgggtgtca ggggaccaca	2040
ggataggaga gaggtatgt taggtacaca cgtagtgtgg gatatcggt tgcaatcgac	2100
aatcgttatg actataccat ggacatcagg cggtacaattt agatataccg atcccgatatac	2160
atatacatca gccggattcc tatcatgttg gtatcagact agtctgtatc tgccaccgc	2220
aactaccggt caagtgtatc tattgtcatt catatccgca tggcccgatt ttaaacttag	2280
attgatgaaa gatacacaaa caatttcgca aacagtcgca ctaaccgaag gggttaggcga	2340
cgaatttagag gaagtgtatag tcgaaaaaac taagcaaaca gtcgcataa ttatccgg	2400
acctaatacatc acacaaaaag tgccatatt gaccgctaaac gaaacaggcg caactatgcc	2460
cgtactgcctc tccgattcaa tcgaaacacg aactacttat atgcattttt acggatccg	2520
aaccgatgtc gaatgtttt taggtaggc cgcatgcgtt cacgttacag agatacagaa	2580

-continued

-continued

agtgtataag gatcttgaga tagacgtatg caatacacca cctagtgggt gtattaacga	4980
tctgttaaaa tcagtcgata gcgaagagat acgcgaatat tgtaaaaaaa aaaaatggat	5040
tatacccgaa ataccgacta atatcgaaag agctatgaat caggcttagt tgattattaa	5100
tactatactt atgttcgtta gtacattggg ttcgtttac gttatataca aattgttgc	5160
acaaacacag ggaccatact caggtaatcc accacataat aagcttaagg cacctacact	5220
tagaccgcgt gtcgttcagg gacctaatac cgaatcgca ctatcactat tgcgtaaaaa	5280
tattatgaca attactacat ctaaaggcga atttacaggg ttagggatac acgatagggt	5340
atgcgttata ccgacacacg ctcacccgg tgatgacgtg ctgggttaacg gtcaaaaaat	5400
taggggtaag gataagtata agttagtcga tccagagaat attaacttag agttaaccgt	5460
actgacactc gatagaaacg aaaaatttag ggatattagg gggtttatat ccgaagatct	5520
tgagggagtc gacgctacat tagtcgtaca ctctaataat tttacgaata ctatactcga	5580
agtccggaccc gttactatgg ccggattgtat taatctatct agtacaccctt ctaatagaat	5640
gattaggatc gattacgcta ctaagacagg gcaatgcggg ggggtgttat ggcgaaccgg	5700
taagatattc ggtatacaccg tagggggtaa cggttagacag gggtttagcg cacaattgaa	5760
aaaacagttat ttctcgaaaa aacagggtca ggttaatcgct agacataagg tttagagat	5820
taatattaat cccgttaata ccgctactaa gtctaaattt catccatcag tgttttacga	5880
cgtattccca ggcgataagg aaccagccgt actatccgtt aacgatccctt gacttgaggt	5940
taagtttaacc gaatcactat tctctaaatgta taagggttaac gttaataccg aacctaccga	6000
aaatatgtta gtcgcagtcg atcattacgc cggtcaattt ctatcactt atataccgac	6060
tagcgaattt acacttaaag aggcaactata cggagtcgac ggactcgaac caatcgat	6120
tactacatca gccggattcc catacggtt tcttaggtttt aaaaaaagggtt atatacttaa	6180
caaagagaca caggataccg aaaaatgaa attttatcta gataaatacg gtatcgatct	6240
accattatgtt acattatatta aagacgaatt gcgatcgtt gataagggtt ggttaggttt	6300
gtcttagattt atcgaagcta gttcacttaa cgatagcggtt aatatgcgtt tgaaatttgg	6360
taatctgtat aaggcatttc accaaaatcc cggagttttaa accggatccg cagtcggat	6420
cgatccagac gtttttttgtt cagtgtatcc atgtctttagt gacggacatc ttatggcatt	6480
cgattactt aatttcgatg catcactatc acccggtttgg ttcgtatgtc ttgagaaagt	6540
gttaactaag tttagggatcg caggatctt tctgataaca tcaatttgcgta atacacatca	6600
tatttttaga gacgaaatat acgttagtcga agggggatcg cctagggat gttccggat	6660
atcaattttt aactctatga ttaacaatat tattattaga acattgtatc ttgacgcata	6720
taagggtatc gatcttagata agcttaagat actcgcatatc ggagacgtt taatcgat	6780
ttatccatatac gaactcgatc cacaagtgtt agcgacattt ggtaagaatt acggatttgc	6840
tattacacctt cccgataaaat ccgaaacttt tacgaaaatg acatgggaga atctgcatt	6900
tctgaaacga tattttaaac ccgatcgtt attccattt ttagtgcattt ccgttatgcc	6960
tatgaaatg atacacgaaat caatttagt gactaaggat caaaaaaaaata cacaggatca	7020
cgttaggtca ctatgtatgt tagcatggca ttcaggcgaa aaagagtata acgaatttcat	7080
aaaaaaaattt agaactaccg atatcggtaa gtgtctgata ttgcccgaat actcagtgt	7140
tagacgttataa tgggttagatc tattcttaggt taacaatata gacacttaat ttgagtagaa	7200
gttaggatcc at	7212

-continued

-continued

gcgtcgatg	gcatattag	gtgatacgcc	gtgggatttt	ggttcgatag	gtggtgtatt	2220
tacgtcggtt	gtaaaattaa	tacatcaa	atttggtagc	gcgtatggtg	tattat	2280
gggtgtatcg	tggacgatga	aaataggta	aggtatatta	ttaacgttgt	taggttaaa	2340
ttcgcgttcg	acgtcgttat	cgtacgtg	tatagcggt	ggtatggta	cgttat	2400
aggtgtatcg	gtacaagcgg	attcgggtt	tgtataaa	tggaaagg	gtgaattaaa	2460
atgtgggtcg	ggtatattt	taacgaatga	agtacatacg	tggacggaac	aatataaatt	2520
tcaagcggat	tcgcccggaa	gttatacgcc	ggcgatagg	aaagcgtgg	aagaagg	2580
atgtggata	cgttcggcga	cgcgtt	aaatataat	tggaaaca	tatcgaatga	2640
attaaatcat	atattattag	aaaatgat	gaaattac	gtatgtat	gtatgtatc	2700
gggtatatta	gcgcaaggta	aaaaatgat	acgtccgca	ccgatgg	ataaaat	2760
gtggaaatcg	tgggtaaag	cggaaata	aggtgcgg	gtacaaaata	cga	2820
aatagatgg	ccgaatacgc	cggaaatgtcc	ggataatcaa	cgtgcgtg	atata	2880
agtagaagat	tatgg	tttg	gtatatttac	gacgaatata	tggtaaaat	2940
gtatacgc	gtatgtatc	atcg	tat	gtcggcgg	ataaaag	2900
acatgcggat	atgggttatt	ggatagaatc	ggaaaaaa	gaaacgtg	gaaat	3060
tgcgtcg	tatagaat	aaacgtgtat	atggccgaa	tgcatacgt	tatgg	3120
tgggttatta	gaatcggaaa	tgataatacc	gaaaatata	ggtgg	tatcgca	3180
taattatcg	cgggttatt	ttacgcaac	ggcggtcc	tggcatt	tagt	3240
attagat	ttt	gat	ttatgt	gtac	gat	3300
tggccgtcg	ttacgtacg	cgcggta	ggtaaaac	atacatg	aat	3360
tccgtgtacg	ttaccgcgt	ta	cg	ttt	atgg	3420
aatacgtccg	gtaaaagaaa	aagaagaaa	tttagt	aa	tcgatgg	3480
gggtgaagta	gattcg	ttt	cgtt	attatgt	tgcataatg	3540
aatgcgttcg	cgttgg	tcg	gtt	aa	gatgg	3600
attaacgat	ggtcaattaa	cgtggat	ttaatacgt	ttatgtataa	tggtaggt	3660
aatgcgtcg	gataaaatgg	gtatgg	tac	gacgtt	ttcg	3720
tatgcgtccg	atgttgcgg	taggttatt	attcgtcg	ttaacgtc	gtga	3780
attattaacg	gtatgtt	at	cg	ttt	ttcg	3840
attaggtat	gtttagc	tg	gg	tata	acgg	3900
geatcaatt	tgg	gg	acgt	tat	ttcg	3960
ttatgcgtgg	aaaacgatgg	cgtat	gat	atcgat	atcg	4020
atcgacgacg	tcgcaaaaaa	cgcgtgg	accgg	tat	ttcg	4080
accgttaacg	atgtttttaa	taacggaaa	aaaatatgg	ggtcgtaa	atcg	4140
aaatgaaggt	ataatggcg	taggtat	atcgat	ttatcg	tat	4200
tgtatgtatcg	ttagcgg	tcgt	ttatcg	ttatcg	ttatcg	4260
atcggttgcg	tccgg	ggatt	tatcg	aaaacg	ggcg	4320
agcggacat	tcgg	gtcg	cgcataat	attatg	atcg	4380
aaaataaaaa	gat	gaga	ac	gtat	gata	4440
agcgatatcg	ggtgtat	atc	gtcg	accgg	gac	4500
aaaaaaaaaa	aaacaacgtt	cgggtgtatt	atggatac	cgtcgcc	cggaa	4560

-continued

acgtgcggta ttagatgatg gtatatatcg tatattacaa cgtggttat tagtcgttc 4620
 gcaagtaggt gtaggtgtat ttcaagaagg tgtatccat acgtatggc atgtaacgcg 4680
 tggtgccgta ttaatgtatc aaggtaaacg tttagaaccg tcgtggcgt cggtaaaaaa 4740
 agatthaata tcgtatggt gtggttggcg tttcaaggt tcgtggatg cgggtgaaga 4800
 agtacaagta atagcggtag aaccggtaa aaatccgaaa aatgtacaaa cggcccccgg 4860
 tacgtttaaa acgcggaaag gtgaagtagg tgcgatagc ttagatttt aaccgggtac 4920
 gtcgggttcg ccgatagtaa atcgtgaagg taaaatagta ggtttatatg gtaatggtgt 4980
 agtaacgacg tcgggtacgt atgtatccgc gatagcgcac gcgaaagcgt cgcaagaagg 5040
 tccggttaccg gaaatagaag atgaagtatt tcgtaaacgt aatttaacga taatggatt 5100
 acatccgggt tcgggtaaaaa cgcgtcgta tttaccggc atagtcgtg aagcgataaaa 5160
 acgtaaatta cgtacgttag tattagcgc gacgcgtgt a tagcgtcgg aaatggcgg 5220
 agcgttaaaa ggtatgccga tacgttataa aacgacggcgt gtaaaatccg aacataacggg 5280
 taaagaataa gtagatttaa tgcgtcatgc gacgtttacg atgcgtttat tgcgggt 5340
 acgtgtaccg aattataata tgataataat ggatgaagcg cattttacgg atccggcgtc 5400
 gatagcggcg cgtgggtata tgcgtcgat tgcgttacg ggtgaagcgg cggcgatatt 5460
 tatgacggcg acgcggccgg gttcggtaga agcgtttccg caatcgaatg cggtaatata 5520
 agatgaagaa cgtgatatac cggAACGTTc gtggattacg gttatgatt ggataacgga 5580
 tttccgggt aaaacggat ggtttgtacc gtcgataaaa tcgggtatg atatagcgaa 5640
 ttgtttacgt aaaaatggta aacgtgttgtt acaattatcg cgtaaaacgt ttgatacgg 5700
 atatcaaaaa acgaaaaata atgattggta ttatgtatg acgacggata tgcggaaat 5760
 gggtgeaat ttgcgtcggt atcgtgtat agatccgcgt cgttggtaa aaccggtaat 5820
 attaaaatgat ggtccggaaac gtcgtatatt agcgggtccg atgcggtaa cggtagcgtc 5880
 ggccggcgcac cgtcggtgc gataggtcg taatcaaat aaagaaggta atcaatata 5940
 atatatgggt caaccgttaa ataatgtatc agatcatgcg cattggacgg aagcgaaaat 6000
 gttattagat aatataaataa cgcggaaagg tataatccg cgttatttg aaccggaaacg 6060
 tggaaatcg gcggcgatag atggatgataa tgcgttacg ggtgaagcgcgt gtaaaacgtt 6120
 tgcgtatattt atgcgtcgat gtcgtgggt tttatgtatg gtcgtataaag tagcgtcgg 6180
 aggtttcaa tattcggtc gtcgtgggt tttatgtatg gtcgtataaag tagcgtcgg 6240
 agaagaaaat atggatgtatc aatatggac gaaagaaggta aacgtaaaat aattacgtcc 6300
 gtcgtgggtt gatgcgcgtc cgttccgta tccgttagcg ttacgtgaat taaagaatt 6360
 tgcggcggtt cgtcggtcg ttcgtgggtt gaaatgtatc aattaccgca 6420
 acatttaacg caacgtgcgc aaaaatcgatc agataattt gtaatgttac ataaatccgaa 6480
 acaagggtggta aacgtatc gtcgtcgat ggaagaatc cccgtatcgt gtaaaacgtt 6540
 aatgttattt gtcgtatc ggttatttac ggggtgggtt gtcgttatttt tttatcgg 6600
 tgcgtgggtt ggtttttttt cgtatgtatc gtcgtatc gtcgttattttt tttatcgg 6660
 atggatggcg tccgtatc gtcgtatc gtcgtatc gtcgtatc gtcgtatc gtcgtatc 6720
 aatgttattt ttaatccgg aacggatcg tcaacgtacg cccgtatcgt gtaaaacgtt 6780
 gtcgtatc gtcgtatc gtcgtatc gtcgtatc gtcgtatc gtcgtatc gtcgtatc 6840
 attagaaacg acgaaaaaaat atttaggtat agtcatgcg gtcgtatc gtcgtatc 6900

-continued

-continued

agtacgtgta caacgtccgg cgaaaaatgg tacggtaatg gatgtatatat cgcgtcgta 9360
 tcaacgttgt tcgggtcaag taggtacgta tggtttaaat acgtttacga atatggaaac 9420
 gcaattaata cgtcaaataatgg aatcggaaagg tatattttcg ccgtcgaaat tagaaacgcc 9480
 gaattttacg gaacgtgtat tagattggtt aaaaaaacat ggtacggaaac gttttaaaacg 9540
 tatggcgata tcgggtgatg attgtgttagt aaaaccgata gatgtacgtt ttgcgacggc 9600
 gttaacggcg ttaaatgata tgggtaaagt acgtaaagat ataccgcaat gggaaaccgtc 9660
 gaaagggttgg aatgattggc aacaagtacc gttttgttcg catcatttc atcaattaat 9720
 aatgaaagat ggtcgtgaaa tagtagtacc gtgtcgtaat caagatgaat tagtaggtcg 9780
 tgcgcgtgta tcgcaagggtg cgggttggtc gttacgtgaa acggcgtgtt taggtaaatc 9840
 gtatgcgcaat atgtggcaat taatgttattt tcatcgctg gatttacgtt tagcggcgaa 9900
 tgcgatatgt tcggcggtac cggtagattt ggtaccgacg tgcgtacga cgtggtcgat 9960
 acatgcgcat catcaatggc tgacgacggc agatatgttac tgcgttatggc atcgtgtatg 10020
 gatagaagaa aatccgtgga tggaaagataa aacgcgtgta tgcgtgtggg aagatgtacc 10080
 gtatttaggt aaacgtgaag atcaatgggt tggttcgtaa atagggtttaa cggcgctgc 10140
 gacgtggcg acgaatatac aagtagcgat aaatcaagta cgtcgtaaa taggtatgaa 10200
 aaattattta gatTTTatgg tgcgtatgg tgcgtatggc aatgaatccg atccggaaagg 10260
 tgcgttatgg taagccaact cattcacaaa ataaaggaaa ataaaaaaatc aaacaaggca 10320
 agaagtcagg cggattaaag ccatagcagc gtaagagcta tgctgectgt gagccccgtc 10380
 caaggacgta aatgaagtc aggccgaaag ccacgggtcg agcaagccgt gctgectgta 10440
 gctccatcggtt ggggatgtaa aaacccggga ggctgcacaaat catggaaatgtt gtagcgtatgg 10500
 ggttagcagac tagtggttag aggagacccc tcccaagaca caacgcacgca gcggggccca 10560
 acaccagggg aagctgtacc ctgggtgtaa ggactagagg ttagaggaga ccccccgcac 10620
 aacaacaaac agcatattga cgctggaga gaccagagat cctgctgtct ctacagcatc 10680
 attccaggca cagaacgcac aaaaatggaa tggtgctgtt gaatcaacag gttct 10735

<210> SEQ ID NO 34
 <211> LENGTH: 10735
 <212> TYPE: DNA
 <213> ORGANISM: Unknown
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 34

agttgttagt ctacgtggac cgacaagaac agtttcaat cggaaagcttgc ttacgttag 60
 ttcttaacagt ttttttttagt agagcagatc tctgtatgaat accacaacgaa aaaaacccgg 120
 acgaccatca ttcaatatgc tttaaacgcgc taggaatagg gtgtcaacccg tatecgcaatt 180
 ggcttaagaga ttctctaagg gactgttaag cggacaggaa ccaatgaaac tggttatggc 240
 attcatagcg tttctgagat ttctcgcaat accgccaacc gcccgaatac tcgctagatg 300
 ggggtcattc aaaaaaaaaacg ggcgaattaa agtgcttagg gggttcaaaa aagagatatc 360
 gaatatgctt aacattatga atagacgtaa gagatccgtt acaatgttat tgcgtatgtt 420
 gccaacccgca ctatgccttc acctaacgcac tagggggggg gaaccacata tgcgtatgt 480
 taagcaagag agaggttaagt cactattgtt caaaacatcc gcccggatgtt atatgtgtac 540
 actaatcgct atggacttagt gcaattgtt cgaagacact atgacatata agtgccttag 600

-continued

gattaccgaa accgaaccag acgacgtcgatgttgggtgt	aacgcacccg aaacatgggt	660
gacatacggaa acatgttcgc aaaccggaga gcatagacgc gataagagat cagtgcact		720
cgcaccacac gtccgattgg ggtagagac tagaaccgaa acatggatgt ctagegaagg		780
cgcataatggaaa cagatacaga aagtcgagac atggccctt agacacccag ggttaccgt		840
aatcgcacta ttctcgac acgcaatcg aacgtcaattt acgcaaaaag ggattatatt		900
catactgctt atgttagtga caccatctat ggctatgaga tgcgtcgaa tcggaaatag		960
ggatccgtt gagggactat ccggagcgac atgggtcgac gtagtgctcg aacacggatc		1020
atgcgttacg actatggcca aagacaacc gacacttgac atagagttac tgaaaaccga		1080
agtactaac ccagccgtac tgagaaagtt gtgtatcgaa gctaagatat cgaatacgac		1140
taccgatagtt aggtgtccaa cgcaaggcga agcgacacta gtcgaagagc aggatacgaa		1200
tttcgttatgt agacgttacat tcgtcgatag ggggtggggaa aacggatgctg gattgttcgg		1260
taaggatca ctgattacat ggcctaaattt caaatgcgtt acgaaaactcg aggtaagat		1320
atgcataac gagaatctga aatatagctt aatcgatcgtt gtcgttacccg gagaccaaca		1380
ccaagtgaaa aacgagacta ccgaacacgg aactaccgca acgattacgc cacaagcccc		1440
tactagcgaa atccaatttga ccgattacgg cgactaaca ctcgattttt cacctagaac		1500
cgattggac tttaacgaaa tgggtctattt gactatgaaa aaaaaatcat ggttagtgca		1560
taagcaatgg ttcttagacc taccactacc atggacttgc ggagcttagt cgtcacagga		1620
gacatggaaat agacaggatc tggtagtgcattt attcaaaacc gcacacgcta aaaaacagga		1680
ggtcgtatgt ttagggtcac aagagggcgc tatgcatacc gcactaacag ggcacccgaa		1740
gatacagact agcggAACGA ctacgatattt cggccggacac cttaaatgtt gactgaaaat		1800
ggacaaaactg atactgaaatg gtatgtcata cgttatgttgc accggatcat tcaaactcgat		1860
aaaagaggcgc gcccggacac aacacggaaat agtggtagtgc caagtgaaaat acgaggaaac		1920
tgcgcacca tggatgatc acgtttcgatc acaggacgaa aaaggcgatc cgacaaacgg		1980
tagactgatt accgcttacc caatcgatc cgataaagag aaaccggatc atatcgatc		2040
cgaaccacca ttccggaaat catatatcgatc agtccggatcc ggagagaaaag cccttaatgt		2100
atcatggttt aaaaagggtt catcaatcgatc taagatgttgc gaagcaaccg ctagaggcgc		2160
tagacgtatgc cccatactgc ggcataccgc atgggatttc ggatcaatcg gaggatgtt		2220
tacgtcagtc ggaaaacttgc tacaccaat ttccggaaacc gcatacgatc tggatgttgc		2280
cgaggatgtca tggactatgc aaatcgaaat cggataactgc ttgacatgttgc tagggatgttgc		2340
ctctagatcg actatgttgc tttatgtatgc tatcgatcgatc ggaatggatc cactataactt		2400
aggcgtaatgc gtgcggccgc atagcgatgc cgttattat tggaaaggaa gagatgttgc		2460
atgcggatcc ggaatattcg ttacgaaatcg agtgcataca tggaccggaaat agtataaatt		2520
ccaaacccatc agtccggaaa gactatctgc cgcataatcgatc aaggcatgggg aggaggaaat		2580
gtgcggaaatc aggtcagcgatc ctagactcgatc aaacattat tggaaacaga tatctaaatcg		2640
actgaatcactt atactgttgc agaacgtatc gaaatgttgc gtcgtatgttgc gagatgttgc		2700
ccggataactc gcacagggtt agaaaatgtt tagaccgaaat cctatggggc ataagtatcg		2760
ttggaaatca tggggaaaat cggaaaattat cggagccgc gtcgtatgttgc ttagatgttgc		2820
aatcgacggaaatc cggataatcgatc cagagtgtcc agacaatcgatc agagcatggaa atacatggaa		2880
ggtcgtatgt ttagggatcc gatgttttgc gactaacata tggatgttgc ttagatgttgc		2940
atatacacaatcgatc atagattgttgc tttatgtatgc tatcgatcgatc ggaatggatc cactataactt		3000

-continued

acacggccat atggggtatt ggatcgaatc cgaaaaaaaaac gagacatgga aactcgctag 3060
agcgtcattc atagaggta agacatgtat ttggcctaag tcacatacat tgtggctaa 3120
cgagggtctt gaggctgaga tgataatccc taagatatac ggccgaccga tatcgcaaca 3180
taactataga ccagggtatt ttacgcaaac agccggacca tggcatctcg gtaagcttga 3240
gttagacttc gatctatgct aaggactac cgtatggtc gacgaacatt gcggtatag 3300
gggacctagt ctgagaacga ctaccgttac cggttaagact atacacgaaat ggtttgttag 3360
gtcatgtaca ctaccaccac ttagattcaa aggccaagac ggatgttggt acggaatgga 3420
gattagacca gtgaaagaga aagagaaaaa cctagtgaaa tctatggtgt cagccggatc 3480
aggcgaagtc gactcatttct cactcggact gttatgcata tcgattatga tcgaagagg 3540
tatgcgtatct agatggtcac gaaaaatgt tatgaccgga acactagccg tttttctgtt 3600
actgactatg gggcaattga catgaaacga tctgatttagg ttgtgttatta tggtcgagc 3660
taacgctagc gataagatgg gtatggaaac gacataactta gcccttatgg ctacatttag 3720
aatgcgacca atgttcgcgcg tagggttact gtttagacgg ttgacttagta gggaaagtgt 3780
attgttaca gtcggactgt cactagtgcg tagcgttgag ttgcctaaact cactcgaaga 3840
gttagggcgcg ggactcgcaaa tggggatttat gatgcttaag ctattgaccg atttccaatc 3900
tcaccaatttgc tgggctacac tattgtcact gacattcggtt aagactacat tctcattgc 3960
ttacgcattgg aaaactatgg ctatgatact gtcaatcggtt agtctattcc cactatgtct 4020
atctacaact agtcagaaaaa ctacatgtt gccagtctta ttgggggtcac taggggtgtaa 4080
accattgact atgtttctga ttaccgaaaa caaaatttgg gggagaaaaatg catggccact 4140
taacgaggga attatggcccg tagggatagt gtcaataactg ctatctagtc tgctaaaaaa 4200
cgacgtgcca ctagccggac cactgatagc cggaggtatg ctaatcgat gttacgtgt 4260
atccggatct agcgccgatc tgcgtactcgaaaagccgaa gaagtgtcat gggagaggg 4320
agccgaacac tctggcgcat cacataacat actagtcgaa gtgcaagacg acggaactat 4380
gaaaattaaaaa gacgaaagaga gagacgatac acttacgata ctgcttaaag cgacactgtt 4440
agcgatatacc ggagtgtatc ctatgtcaat acccgctaca ctattcggtt ggtatffff 4500
gaaaaaaaaaa aaacagagat ccggagtgtt gtgggatatac cctagtcac ccgaactcg 4560
gagagccgtt ctcgacgacg gaatatatac gatactgca cggggattgc tcggtaggtc 4620
acagggtcgaa gtggggatgtt ttccaggaggg agtgtttcac actatgtggc acgttacgag 4680
aggcgccgtt ctgatgtatc agggtaagag actcgaaacct agttggctta gctttaaaaaa 4740
agacctaatac tcatacggag ggggggtggag atttcaggggg tcatggaaacg ccggagaggg 4800
agtgcataatg atagccgtcg aacccggtaa gaaccctaaa aacgtgcaaa ccgcacccgg 4860
aacgtttaag acacccgaaag gcgaaatggg cgcacatcgat ctcgacttta aacccgaaac 4920
tagcgatca ccaatcgatca atagagaggg taagatagtc ggattgtacg gaaacggagt 4980
ggttacgact agcggaaacat acgttagcgca aatagcgcaaa gegaaagcgt cacaagaggg 5040
accactacca gagatagagg acgaaatgttt tagaaagegt aatctgacaa ttatggaccc 5100
acatccccggaa tccggtaaga ctagacgtt ctcgtccagca atcgatgggg aggcaattaa 5160
gagaaatgtt cgtacacttag tggatggcccc aactagatgt gtcgttagcgaaatggccg 5220
agcccttaag ggaatgccaa ttaggtatca gactaccggc gttaaatgtccg aacataccgg 5280
taaggagata gtcgacccat ttaggtatca gactaccggc gttaaatgtccg aacataccgg 5340

-continued

gagagtgcct aactataata tgataattat ggacgaagcc cattttacag acccagcctc	5400
aatcgccgct agggggata tttagtactag ggtggaaatg ggcgaagccg ccgcaatttt	5460
tatgacagcg acaccacccg gatccgtcga a cggttccg caatctaacg ccgttataca	5520
ggacgaagag agagacatac ccgaacggtc atgaaatagc ggataacgatt ggataaccga	5580
tttcccggt aagacagtgt gttcgttacg gtcaattaag tccggaaacg atatcgctaa	5640
ttgtcttaga aaaaacggta agagagtggt gcaattgtct agaaagacat tcgataccga	5700
ataccaaaaa actaagaata acgattggga ttacgtatgt acaaccgaca tatccgaaat	5760
gggcgctaac tttagagccg atagggtat agaccctaga cgggtctga aaccagtgtat	5820
actgaaagac ggaccagaga gagtgatact agccggacca atgcccgtga cagtcgcac	5880
cggccgacaa cgtagggac ggatagggag aaacccaaac aaagagggag accaatacat	5940
atatatggga cagccactga ataacgacga agaccacgca cattggaccc aagcgaaaat	6000
gctttagac aatattaaca cacccgaagg gattataccc gcactattcg aacccgaacg	6060
cggaaaaatcc gccgcaatcg acggcgaata tagactgaga ggcgaagctt gaaagacatt	6120
cgtttagctt atgcgttagag gcgatctacc cgtatgggtt tcatacaaaag tcgctagcga	6180
agggtttcag tattccgata ggagatgggtt ttccgacggc gaacggaaata accaagtgt	6240
cgaagagaat atggacgttg agatatggac taaaagagggc gaacgaaaaa aactgagacc	6300
tagatggta gacgcacgta catattccga tccgttagcc cttagagagt taaaagagtt	6360
cgcagccgtt aggagatccg tttagcgaga tctgatactc gaaatcggtt agctaccaca	6420
acacctaacg caacgagccc aaaacgcact agacaatctg gttatgttc ataactccga	6480
acagggcggtt aaggcatata gacacgctat ggaagagtta cccgatacaa tcgaaacact	6540
atatcgactc gcactgtatcgccgacttgcgatccgacttgcgacttgcgatccg	6600
taggggggtta ggttaagacat caatcgact gttatgcgtt atcgcatctt ggcactatt	6660
gtggatggctt agcgtcgaac cacattggat agccgcatctt attatactcg aatttttctt	6720
tatgggttta ctgatacccg aacccgatag gcaacgaaca ccacaggata accaattggc	6780
atacgtcgta atcggtactat tggttatgtt actgacatgcgatccgacttgcgatccg	6840
gtcgagaca acgaaaaaaag acctaggat cggacacgccc gcagccgaaa atcaccatca	6900
cggccgcaatg ttagacgtcgatccgacttgcgatccgacttgcgatccgacttgcgatccg	6960
aacgacaattt attacgccta tggatgcgaca tacaatcgag aataactaccg ctaacatatc	7020
gctaaccgctt atagcgtatccgacttgcgatccgacttgcgatccgacttgcgatccg	7080
tagtaagatg gatatcggtt tgccactgtt agccttaggg tggttatgtt aggttaaccc	7140
attgacattt accgcgtccg tactttatgtt agtgcgtatccgacttgcgatccgacttgcgatccg	7200
actgcaagcg aaagcgacac gcaagcgcaaaaagaacc gcagccggaa ttatgaaaaa	7260
ccctacatgcgatccgacttgcgatccgacttgcgatccgacttgcgatccgacttgcgatccg	7320
cgaaaaacag ttgggacaga ttatgttactt gatactgtgtt acgtcacaga tactgttat	7380
gcgaaactaca tgggactat gcaatcaat tacactcgca accggaccat tgactacatt	7440
gtggggagggtt tcacccggta agttttggaa tacgactata ggcgttagta tggctaaat	7500
ttttaggggg tcatatcttgcgatccgacttgcgatccgacttgcgatccgacttgcgatccg	7560
agggggacgtt agggggacat ggcgtccgacttgcgatccgacttgcgatccgacttgcgatccg	7620
attgaaccaa ttgtctaaat ccgaattttaa cacttataag agatccggaa taatcgaaat	7680
cgatagatcc gaaacgaaatg agggactgaa acgcggagag acaacgaaac acggccgttag	7740

-continued

taggggaacc gctaagctta gatggttcgt tgagcgtaac ttagtcaaac cagaggtaa 7800
 ggttaatcgat ctccgatgcg gtggggggg gtggtcatac tattgcgcg gactgaaaaa 7860
 agtgaccgaa gtgaaagggt atactaaggg cgAACCCGGA cacgaagagc caataccgat 7920
 ggcaacatac ggatgaaacc tagtggaaact gtattccggt aaggacgtat tctttacacc 7980
 acccgaaaaa tgcgatacac tggatgcga tataggcgaa tctagtcata accctacaat 8040
 cgaagaggaa cgtacactta gagtgcttaa aatggtcgaa ccatggtga gggggatca 8100
 gttttgtatt aagatactga atccatatac gcctagcgta gtcgagacac tcgaacagat 8160
 gcaacgaaaa cacggcgaa tgcttagttag aaacccacta tcttagaaat ccacacacga 8220
 aatgtattgg gtgtcatgcg gaaccggaaa catagtgtca gccgttaata tgacttagtag 8280
 aatgtactg aatcggtta caatggccca tagaaaaccg acatacgaaa gagacgtcga 8340
 cctaggagcc ggaactagac acgttgcgt cgaacccgaa gtcgctaact tagacataat 8400
 cggacagaga atcgagaata ttAAAACGAA acacaaatcg acatggcatt acgacgaaga 8460
 caatccgtat aagacatggg catatcacgg atcatacgaa gtgaaaccta gcccgtccgc 8520
 tagttctatg gttaacggag tggtagact gttgactaag ccatggacg tgataccgat 8580
 ggtgacacaa atcgctatga ccgatacgac accattcgaa cagcaagag tgttaaaga 8640
 gaaaggcgc actagaacac ctaaagcggaa acgcggaaacc gcacagatgg tggaggtac 8700
 cgctagatgg ttgtgggggt tcctatctag aaacaaaaaa cctaggatt gtacacgcga 8760
 agagttaca cgaaaagtga gatctaacgc cgcaatcgaa ggcgtattcg tagacgaaaa 8820
 ccaatggaaat tccgtaaag aggccgtcgaa agacgaacgg ttttggact tagtgcata 8880
 ggaacgcgaa ttgcataaac agggtaaatcg cgcaacatgc gtatataata tgatggtaa 8940
 ggcgaaaaa aagttaggcg aattcggtaa ggctaaaggaa tctagagcgaa tatggtat 9000
 gtgggttaggc gctaggttc tcgaaatcgaa agcactcgaa tttatgcacg aagaccattg 9060
 gttttctaga gagaatagtc tatccggagt cgagggagag ggattgcata aactcgata 9120
 tataactgaga gacatatacta agataccgg agggaatatg tacgcgcacg atacagccgg 9180
 atgggataca cgaattaccg aagacgatct gcaaaacgaa gcaaaaattt ccgatattat 9240
 ggaacccgaa cacgcactgt tagegacatc tatttttaa tggacatatc agaataagg 9300
 cgtagggtg caacgacccg ctaaaaacgg aacggttatg gacgtgatata gtaggagaga 9360
 ccaacgcgaa tccggacagg tcggaacata cggactgaat acgtttacga atatggaaac 9420
 gcaattgatt agacagatgg agtctgaggg gatattctca cctagcgaac tcgagacacc 9480
 taacctagcc gaacgggtac tcgattgggtt gaaaaaacac ggaaccgaaa gactgaaaag 9540
 aatggcaatt agcggagacg attcgtagt gaaaccaatc gacgatagat tcgcaaccgc 9600
 actaaccgc cttaaacgata tgggtaaaggat gagaaaagac ataccgcaat gggagccatc 9660
 taagggatgg aacgattggc aacaggtgcc attttgcata catcacttcc accaattgat 9720
 tatgaaagac ggttagggaga tagtcgtgcc atgttaggaat caggacgaac tcgttaggt 9780
 ggcttagatgc tcacaggag ccggatggtc acttagggaa accgcatgcc taggtatgc 9840
 atacgctaa atgtggcaat tggatgtatccatcgacg gatctgagac tagccgtaa 9900
 cgctatatacg tctgcgtac cagtcgattt ggtgccaact agtggaaacgaa catggtctat 9960
 acacgcacat caccaatggaa tgacaaccgaa agacatgcttccgtttggaa atagggtgt 10020
 gatcgaaagag aatccatggaa tggaggataa gacacacgtg tctagttggg aggacgtgcc 10080

-continued

```

atatactcgga aaacgcgaag accaatggtg cgatcacta atcgattga ccgctagagc 10140
gacatgggc actaacatac aggtcgcaat taaccaatg agacgattga tcggaaacga 10200
gaattatctc gacttatga catctatcaa aagattcaaa aacgaatccg atcccaagg 10260
cgcaactatgg taagccact cattcacaaa ataaaggaaa ataaaaatc aaacaaggca 10320
agaagtccagg ccggattaaag ccatagcagc gtaagagcta tgctgcctgt gagccccgtc 10380
caaggacgta aatgaagtc agggccaaag ccacggttcg agcaagccgt gctgcctgt 10440
gtccatcggt gggatgtaa aaaccggga ggctgcaaac catggaagct gtacgatgg 10500
ggtagcagac tagtggtagt aggagacccc tcccaagaca caacgcagca gccccccca 10560
acaccaggaa aagctgtacc ctggtggtaa ggactagagg ttagaggaga ccccccgcac 10620
aacaacaac acatattga cgctggaga gaccagat cctgcgtct ctacagcatc 10680
attccaggca cagaacgcca aaaaatggaa tggtgctgtt gaatcaacag gttct 10735

```

<210> SEQ ID NO 35

<211> LENGTH: 9181

<212> TYPE: DNA

<213> ORGANISM: Human immunodeficiency virus type 1

<400> SEQUENCE: 35

```

ggctctctcg gttagaccag atctgagcct gggagctctc tggctaacta gggAACCCAC 60
tgcttaagcc tcaataaaagc ttgcctttag tgcttcaagt agtgtgtgcc cgtctgttgt 120
gtgactctgg taactagaga tccctcagac ccttttagtc agtgtggaaa atctctagca 180
gtggcgcccg aacagggacc taaaagcga aggaaaccga gaggagctct ctcgacgcag 240
gactcggctt gctgaagcgc gcacggcaag aggcgagggg cggcgactgg tgagtacgcc 300
aaaaattttg actagcggag gctagaagga gagagatggg tgcgcggtcg tcggatttat 360
cggttgggtga attagatcgt tggaaaaaa tacgtttacg tccgggttgt aaaaaaaaaat 420
ataaattaaa acatatagtt tggcgctcgt gtgaattaga acgttttgcg gttaatccgg 480
gtttattaga aacgtcgaa gggttcgtc aaatattagg tcaattacaa ccgtcggtac 540
aaacgggttc ggaagaatta cgttcggttataatacggt agcgacgttataattgtgtac 600
atcaacgtat agaaataaaa gatacgaaag aacgtttaga taaaatagaa gaagaacaaa 660
ataaattcgaa aaaaaaaagcg caacaagcgg cggcgatatac gggtcattcg aatcaagtat 720
cgcaaaatatac tccgatagta caaaatatac aaggtaat ggtacatcaa gcgatatcgc 780
cgcgtaatggg aatgcgtgg gtaaaatgg tagaagaaaa agcggtttcg ccggaaatgg 840
taccgtt ttcggcggtt tccggatgtcg aacgtttagt agatgttacgatgtt 900
atacggtagg tggatcatcaa cggcgatgc aaatgttaaa agaaacgata aatgaagaag 960
cgccggaaatg ggatcggttta catccggatgc atgcgggtcc gatagcgccg ggtcaatgc 1020
gtgaaccgcg tggatcggtt atacgggttgcg acgttgcgtt gttacaagaa caaatagggtt 1080
ggatgacgaa taatccggccg ataccggtagt gtgaaatata taaaatggg ataataatgg 1140
gtttaaataa aatgttacgt atgttacgcg cgacgttgcgtt attagatata cgtcaaggtc 1200
cgaaagaacc gtttcgtgtat tatgttagatc gttttataa aacgttacgt gcgaaacaaag 1260
cgtcgcaaga agtaaaaaat tggatgacgg aaacgttatt agtacaaaat gcgatccgg 1320
attgttaaaaac gatattaaaac gctttagtgc cggcgccgcgtt gttacaagaa atgttgcgg 1380
cggttcaagg tggatgttgtt ccgggttataa aacgttgcgtt attagcgaa gcgatgtcgc 1440
aagtaacgaa ttcggcgacg ataatgttgc aacgttgcgtt ttttgcgtt aacgttaaaa 1500

```

-continued

tagtaaaatg tttaattgt ggtaaagaag gtcatacggc gcgttaattgt cgtgcgcgc	1560
gtaaaaaagg ttgttgaaa tggtaaaag aaggtcatca aatgaaagat tgtacggAAC	1620
gtcaagcgaa ttttttaggg aagatctggc cttcctacaa gggaggcca gggatTTc	1680
tccagagcg accagagCCA acagccccac cagaagagAG ctccaggtct ggggtagaga	1740
caacaactcc ccctcagaag caggagCCGA tagacaagGA actgtatccc ttaacttccc	1800
tcaaggtaact ctggcaac gaccctcgT cacaataaAG atagggtgtc aataaaAGA	1860
agcgTTatta gatacgggtg cggtatGATAc ggtattAGAA gaaatgtcgt tacGGGTcG	1920
ttggaaaccg aaaatgatAG gtggatAGG tggtttata aaagtacgtc aatATGATCA	1980
aatattaata gaaatATGTG gtcataaAGC gataggtaCG gtattAGTAG gtcgcACGCC	2040
ggtaaatata ataggTCgtA attttaaAC gcaaataAGT tgtaCgttaA atttccGAT	2100
atcgccgata gaaacggTAC cggtAAAATT AAAACCGGGT atggatggTC cgaaAGTA	2160
acaatggccg ttaacggAAAG AAAAATAAA AGCgttagta gaaatATGTA CGGAAATGGA	2220
aaaAGAAGGT AAAATATCGA AAATAGGTCC ggAAAtCCG tataatacGC CGGTATTGc	2280
gataaaaaaa aaAGATTcGA CGAAATGGCG taaattAGTA gatTTcGTG aatTAATAA	2340
acgtacgcaa gatTTTGGG aagtacaATT aggtataCCG catCCGGCGG gttttAAAAA	2400
aaaaaaaaatcG gtaacggTAT tagatgtAGG tgatgcgtAT tttcggTAC CGTTAGATGA	2460
agatTTcGT aaatatacGG CGTTACGAT ACCGTcGATA AATAATGAA CGCCGGTAT	2520
acgttatcaA tataatgtat tacGcaagg ttggAAAGGT tcGCCGGCGA tatttcaATC	2580
gtcgatgacg AAAATATTAG AACCGTTcG taaacAAAT CGGATATAG TAATATATCA	2640
atatatggat gatttatATG taggttcGA tttAGAAATA ggtcaacATC gtacgAAAT	2700
agaagaatta cgtcaacATT tattacgttG gggTTAACG acGCCGGATA AAAAACATCA	2760
aaaAGAACCG CGTttttat ggtgggttA tgaattACAT CGGATAAAT ggacggTACA	2820
accgatAGTA ttaCCGGAAA aagattcgtG gacggtaat GATATAACAA ATTAGTAGG	2880
taaattAAAT tgggegtcgc aaatatacC gggTATAAA GTACGTCAAT TATGAAATT	2940
attacgttgt acgaaAGCgt taacggAAgt aataccgttA acggAAAGAG CGGAATTAGA	3000
attageggAA aatcgtAAA tattAAAGA accggTACAT ggtgtatATT atgtccgtc	3060
gaaAGATTa atacggAAA tacAAAACA aggtcaAGGT caatggacGT atcaaATA	3120
tcaagaACCG tttAAATTAA taaaACGGG taaatATGcG cgtatgcgtG gtgcgcatac	3180
gaatgtgtA aaacaattAA CGGAAGCGGT acaaaaATA ACgAcGGAAat CGATAGTA	3240
atggggtaAA acGCCGAAAT ttaaattACC gataAAAAA gaaACGTGGG aaACGTGGT	3300
gacggAATAT tggcaAGCGA CGTGGATAcc ggaatGGGAa tttgtAAata CGCCGCCGTT	3360
agtaAAATTa tggTATCAAT tagAAAAGA accgatAGTA ggtgcggAAA CGTTTATGT	3420
agatggTcGc gCGAATCGT AAACGAAATT aggtAAAGCG ggttATGTAA CGAATCGTGG	3480
tCGTCAAATAA gtagtaACGT taacggatac gacgaatCAA AAAACGGAAT tacaAGCGAT	3540
atatttagcG ttacaAGATT CGGGTTAGA agtAAATAA gtaacGGGT CGCAATATGC	3600
gttaggtata atacaAGCGC AACCGGATCA ATCGGAATCG gaattAGTAa ATCAATAAAT	3660
agaacaATTa ataaaaAAAG AAAAAGTATA tttAGCGTGG gtaccGGCGC ataaAGGTAT	3720
aggtggtaat gaacaAGTAG ataaATTAGT ATCGGCGGGT ATACGTAAAG TATTATTT	3780
agatggtata gataaAGCGC aagatGAACA tggAAAATAT CATTGAAATT GGCGTGCAT	3840

-continued

ggcgtcgat tttaatttac cgccggtagt agcgaaagaa atagtagcgt cgtgtataa	3900
atgtcaatta aaaggtaag cgatgcattgg tcaagtagat tgttcgccgg gtatggca	3960
attagattgt acgcatttag aaggtaaagt aatattagta gcggtacatg tagcgtcggg	4020
ttatatagaa gcggagtaa tacccggcga aacgggtcaa gaaacggcgt atttttatt	4080
aaaatttagcg ggtcggttgcg cggtaaaaac gatacatacg gataatggtt cgaatttac	4140
gggtgegacg gtacgtcggtt cgtgttgggt ggccgggtata aaacaagaat ttggatacc	4200
gtataatccg caatcgcaag gtgttagtata atcgatgaat aaagaattaa aaaaataat	4260
aggtcaagta cgtgatcaag cggAACATT AAAACGGCG GTACAAATGG CGGTATTAT	4320
acataatttt aaacgtaaag gtggtatagg tggttattcg gcgggtgaac gtatgtaga	4380
tataatagcg acggatatac aaacgaaaga attacaaaaa caaataacga aaataaaaaa	4440
tttcgtgta tattatcgta attcgcgtaa tccgttatgg aaaggccgg cgaaattatt	4500
atggaaaggt gaagggtcggt tagtaataca agataattcg gatataaaag tagtaccgc	4560
tgcgtaaacg aaaataatac gtgattatgg aaaacagatg gcaggtgatg attgtgttgc	4620
aagtagacg gatgaggatt agaacgtgga aatcgtagt aaaacatcat atgtatgtat	4680
cgggtaaacg gcgtgggtgg ttttacgtc atcattatga atcgccgc catcgctatat	4740
cgtcggaaatg acatataccg ttaggtgatg cgcgtttatg aataacgacg tattgggtt	4800
tacatacggg tgaacgtgat tggcatatgg gtcaagggtgt atcgatagaa tggcgtaaaa	4860
aacgttattc gacgcgtatg gatccggat tagcggatca attaatacat ttatattatt	4920
ttgattgttt ttcggattcg gcgatacgta aagcgttatt aggtcatata gtatcgccgc	4980
gttgcgtata tcaagccgggt cataataaaag tagttcgat acaatattta gcgttagcgg	5040
cgttaataac gccgaaaaaa ataaaaccgc cgttaccgtc ggtaacgaaa ttaacgaaag	5100
atagatggaa caagccccag aagaccaagg gccacagagg gagccacaca atgaatggac	5160
actagagctt ttagaggagc ttaagaatga agctgtttaga cattttcttggatggct	5220
ccatggctta gggcaacata tctatgaaac ttatgggat acttgggcag gagtgaaagc	5280
cataataaga attctgcaac aactgctttt tatccatttt cagaattggg tgcgtacata	5340
gcagaatagg cgttactcga cagaggagag caagaaatgg agccgtatg tccttagacta	5400
gagccctggaa agcatccagg aagtcagctt aaaaactgtttt gtaccaatgg ctattgtaaa	5460
aagtgttgc ttcattgcca agttgttcc ataacaaaag ctttaggc catctatggc	5520
aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagcttct	5580
ctatcaaagc agtaagtagt acatgtatg caacctatac caatgtatgc aatgtatgc	5640
tttagtagtag caataataat agcaatagtt gtgtggtcca tagtaatcat agaatatagg	5700
aaaatattaa gacaaagaaa aatagacagg ttaattgata gactaataga aagagcagaa	5760
gacagtggca atgagagtga aggagaaata tcagcacttg tggagatggg ggtggagat	5820
gggcaccatg ctccttggga tggatgtatgt ctgtgtcg acggaaaaat tatgggtaac	5880
ggatattatgtt ggtgtaccgg tatggaaaga agcgacgacg acgttatttt gtgcgtcgg	5940
tgcgtaaacgc tatgatacgaa aagtacataa tgtatggcgt acgcgtatgcgt gtgtaccgc	6000
ggatccgaat ccgcgtaaag tagtattatgtt aatgtatgc gaaaattttt atatgtggaa	6060
aaatgtatgtt gtagaaacaaa tgcgtatgaa tataatgtatgc ttatggatc aatgttaaa	6120
accgtgtgtat aatgtatgc cgttattgtt atcgatggatt taaaatgtaa	6180
tacgtatgtat aatgtatgc cgttattgtt atcgatggatt taaaatgtaa	6240

-continued

-continued

cgtataaagc ggccgttagat ttatgcatt tttaaaaga aaaagggttgtt tagaaaggtt	8640
taatacaccc gcaacgtcgta caagatataat tagatttatg gatatcatat acgcggatgtt	8700
atttccgga ttaacaaaat tatacgccgg gtccgggtgt acgttatccg ttaacgtttg	8760
gttgggttta taaatttagta ccggtagaac ccggataaaat agaagaagcg aataaagggt	8820
aaaatacgtc gttattacat ccggtagtcgta tacatggat ggatgtatccg gaacgtgaag	8880
tattagaatg gcgttttgat tcgcgttttag cggttcatca tgtagcgcgtt gaattacatc	8940
cggaatattt taaaaattgt taacatcgag ctgtctacaa gggactttcc gctggggact	9000
ttccaggagg gcgtggcctg ggcgggactg gggagtgcc agccctcaga tcctgcata	9060
aaggcagctgc ttttgcctg tactgggtct ctctgggttag accagatctg agcctgggag	9120
ctctctggct aactaggaa cccactgctt aagcctcaat aaagcttgc ttgagtgtt	9180
c	9181

<210> SEQ_ID NO 36
<211> LENGTH: 9181
<212> TYPE: DNA
<213> ORGANISM: unknown
<220> FEATURE:
<223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 36

ggtcctctcg gtttagaccag atctgagccct gggagctctc tggctaaacta gggAACCCAC	60
tgcttaagcc tcaataaagc ttgccttgag tgcttcaagt agtgtgtgcc cgtctgttgt	120
gtgactctgg taactagaga tccctcagac ccttttagtc agtgtggaaa atctctagca	180
gtggcgcccc aacagggacc tgaaagcgaa agggaaacca gaggagctct ctcgacgcag	240
gactcggctt gctgaagcgc gcacggcaag aggccgggg cggcgactgg tgagtacgcc	300
aaaaattttg actagcgag gctagaagga gagagatggg agctagagct agcgtactta	360
gcggaggcga actcgataga tggaaaaaaa ttagactgag accaggggga aaaaaaaaaat	420
ataaaactgaa acatatacgta tgggctagta gagagttaga gagattcgca gtaaaaccag	480
ggttactcga aactagcgaa ggggttagac agatactggg acaattgcaa cctagtctgc	540
aaacccggatc cgaagagctt agatcaactat ataatacagt cgcaacacta tattgtgtac	600
accaacgaat cgaaattaag gatacgaaag aggcatcaga caaaatcgaa gaggaacaga	660
ataagtctaa aaaaaaagcg caacaggcag cagccgatac cggacatagt aatcaggat	720
cgcggaaaatccaatcgta cagaatatac agggacaat ggtacccaa gcgatatacc	780
ctagaacact taacgcattt gtttaggtat tcgaagagaa agcatttagt ccagaggtaa	840
tacctatgtt tagcgcattt agcgaaggcg caacaccaca ggatctgaat actatgttta	900
ataccgtagg ggggcataa gccgcaatgc aaatgcttaa agagacaatt aacgaagagg	960
cagccgaaatg ggatagagtg catcccgat acgcaggacc aatcgacca ggacaaatga	1020
gagaacctag gggatccgat atgccggaa ctactagtac attgcaggaa cagatagggt	1080
ggatgactaa taatcccacct ataccgtat ggcggatata caaaatcgaa ataatactgg	1140
gactgaataa gatagttaga atgtatagtc caacttagtat actcgatatt agacaggac	1200
ctaaagaacc tttagggat tacgttagata gattctataa aacacttaga gccgaacagg	1260
ctagtcaaga ggttaagaat tggatgacag agacactatt agtgcggaa gctaatcccg	1320
attgtaaaac tatacttaag gcactaggac cagcagcaac actcgaaagaa atgatgacag	1380
catgtcaggg agtaggggaa ccaggacata aagctagagt gttagccgaa gctatgtcac	1440

-continued

aggttaacgaa tagcgcaaca attatgatgc aaagaggtaa ttttagaaac caacgaaaaa	1500
tgcgttaagtg tttaattgc gggaaaaggagg gacataccgc tagaaaattgt agggcaccta	1560
gaaaaaaaaagg gtgttggaaa tgcgaaaaag agggacatca aatgaaagat tgtaccgaaa	1620
gacaggctaa ttttttaggg aagatctggc cttcctacaaa gggaaaggcca gggaaatttc	1680
ttcagagcag accagagcca acagccccac cagaagagag cttcaggtct ggggttagaga	1740
caacaactcc ccctcagaag caggagccga tagacaagga actgtatcct ttaacttccc	1800
tcaggtcact ttggcaac gaccctctgt cacaataaag atagggggac aattgaaaga	1860
ggcactatta gacacaggag cagacgatac agtactcgaa gaaatgtcac taccaggttag	1920
atggaaacct aaaatgatag gggggatagg ggggtttata aaggtaagac aatacgatca	1980
gatactaattc gaaatttgcg gacataaagc aatcggaaaca gtgttagtgc gacctacacc	2040
cgttaaacata atcggttagga atctgttaac acagataggg tgtacattga atttccaat	2100
tagtccaaatc gaaaccgtac ccgtaaaact gaaaccaggat atggacggac ctaaggttaa	2160
gcaatggcca ttaaccgaag agaaaattaa ggcatttagtc gaaatttgcg ccgaaatggaa	2220
aaaagagggt aagatttagta agataggacc agagaatccaa tataatacac ccgttattcgc	2280
aattaaaaaa aaagatagta cgaaatggag aaagtttagtc gatTTtagag agttaaaca	2340
aagaacacaa gacttttggg aggtacagtt agggataccaa catccagcag gactgaaaaa	2400
aaaaaaaaagc gtaaccgtat tagacgttagg agacgcatac tttagcgtac cattagacga	2460
agacttttaga aagtataccg catttacgt acctagtatt aataacgaaa caccaggaaat	2520
taggtatcaa tataacgtac tgccacaggatggg atggaaagggt tcaccagcta tattccaaatc	2580
tagtatgact aagataactcg aaccttttag aaaaacagaat ccagacatag tgatatacca	2640
atatatggac gatctatacg taggttagcgtatc tctagagata gggcaacata gaacgaaaat	2700
cgaagagctt agacagcatc tggttagatg gggattgtact acacctgaca aaaaacacca	2760
aaaagaacca ccattcttat ggtatggata cgaatttgcgtatcccataat ggacagtgcata	2820
acctatagtg ttacccgaaa aagactcatg gacagtaaac gatataacaa aactggtagg	2880
taagttgaat tgggttagtc agatataccaa cggaaatTAAG gttaggcaat tggtaaact	2940
gttaaggggaa actaaggcac taaccgaagt gataccgtt accgaagagg cagagtttaga	3000
gttagccgaa aatagagaga tacttaagga acccgatcac ggagtgtatt acgtatccat	3060
taaggatctg atagccgaaa tacaaaaaca gggacaggaa caatggacat atcagatata	3120
tcaggaacct ttAAAAAATC tgaaaaccgg taagtacgtc agaatgaggg gagcacatac	3180
aaacgcgtt aagcaattga ccgaaggccgt aaaaaaaatt acaaccgaat caatcgtaat	3240
ttggggtaag acacctaaat ttaatttgc tataaaaaaa gagacatggg aaacatgttg	3300
gactggtatg tggcaagcta catggatacc cgaatggaaatc ttcgtaaata caccaccatt	3360
agtggaaatttggt ggttatcaac tcgaaaaaaa accaatcgta ggagccgaaa cattttacgt	3420
agacggagca gctaataggg aaactaagtt aggttaaggca ggatacgatcc cgaataggg	3480
gagacagaaa gtcgttacac taaccgatac aactaaccaa aaaaccgaac tgcaagcgat	3540
atacttagcg ttacaggata gcggattaga ggttaaacata gtgacagact cacaatacgc	3600
attagggatt atacaggcac aacccgatca atccgaaagc gaaatgtta atcagataat	3660
cgaacaatttgc attaaaaaaag agaaatgtta tcttagcatgg gtaccagcac ataaggaaat	3720
ggggggaaac gaacaggttagt ataagttgt tagcgccaggaa attagaaagg tactgtttt	3780

-continued

agacggaata gacaaagcgc aagacgaaca cgaaaaatata catagtaatt ggagagctat	3840
ggcttagcgat ttaatctac caccctgtat cgctaaagag atagtcgcat catgtataaa	3900
gtgtcaattt aaaggcgaag ctatgcacgg acaggctcgat tgttagtccag ggatatggca	3960
atggattgtt acacacttag agggtaaggt aatactagtc gcagtacacg tagctacgg	4020
atataatcgaa gccgaagtga taccagccga aaccggacag gaaaccgcat actttctgtt	4080
aaagtttagcc ggtatgtgc cagttaaagac tatacatacc gataacggat ctaatttac	4140
aggagcaacc gttagggcag catgttggtg ggcaggaaat aaacaggaat tcggaatacc	4200
atataatcca caatcacagg gagtagtcga atctatgaat aaagagttaa aaaaaatttat	4260
cggacagggtt agagaccaag ccgaacacct taaaaccgca gtgcaaatgg ccgtattcat	4320
acataatttt aaaagaaaaag gggggatagg ggggtatacg gcaggcgaaa gaatcgtaga	4380
cataatcgca actgacatac agactaaaga gttacagaaa cagattacta agatacagaa	4440
ttttagatgtt tattatagag actctagaaa tccattatgg aaaggaccag ctaaactgtt	4500
atggaaaggc gaaggcgcag tagtgatatac ggataatagc gatattaagg tagtgcctag	4560
acgaaaagcg aaaatttata gggattatgg aaaacagatg gcaggtgtatg attgtgtggc	4620
aagttagacag gatgaggatt agaacatggaa aatcaactatg taagcatcat atgtacgtt	4680
gcggtaaggc tagggatgg ttttatagac atcattacga atcaccacat cctagaatat	4740
ctagcgaagt gcatataccg tttaggcgacg ctagattatg gataactaca tattggggat	4800
tgcatacagg cgaaagggtt tggcatctag gacagggagt gtcaatcgaa tggcgaaaaa	4860
aaagatatacg tacacaggta gacctagatg tagccgtatc attgatacac ctatatttt	4920
tcgattgttt tagcgatagc gcaatttagaa aggcaactatt agggcatata gtgtcaccta	4980
gatgcgaata tcaagccggc cataataagg tagggtaact gcaatatacg gcattagccg	5040
cactaattac acctaaaaaa attaaaccac cactacctag cgtaactaaat ttaaccgaaag	5100
acagatggaa caagccccag aagaccaagg gcccacaggg gagccacaca atgaatggac	5160
actagagctt ttagaggagc ttaagaatga agctgtttaga cattttctta ggatttgct	5220
ccatggctta gggcaacata tctatgaaac ttatggggat acttgggcag gagtggaaagc	5280
cataataaga attctgcaac aactgctgtt tatccatttt cagaattggg tgcacata	5340
gcagaatagg cgttactcga cagaggagag caagaaatgg agccagtaga tccttagacta	5400
gagccctgga agcatccagg aagtcagcct aaaactgtttt gtaccaatgt ctattgtaaa	5460
aagtgttgc ttcattgcca agtttttttataacaaatg ctttaggcatttcctatggc	5520
aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagcttct	5580
ctatcaaagc agtaaagtatg acatgtatgc caacctatac caatagtagc aatagtagca	5640
ttagtagtag caataataat agcaatagtt gtgtggtcca tagtaatcat agaatatagg	5700
aaaatattaa gacaaagaaa aatagacagg ttaattgtata gactaataga aagacgagaa	5760
gacagtggca atgagagtgaa agggaaaata tcagcacttg tggagatggg ggtggagatg	5820
gggcaccatg ctccctggga tggtgtatgtatgtgc accgaaaaat tgggggttac	5880
ggtatattac ggagtaccccg tatggaaaga ggcaactact acactattttt gcgctacgca	5940
cgctaaagca tacgataccg aagtgcataa cgtatggct acacacgcat gcgtacccat	6000
cgatcctaatac ccacaagagg tagtgttagt taacgttaacc gaaaatttttataatgtggaa	6060
aaacgatatacg gtcgaacaaa tgcacgaaga cataattacg ttatggacc aatcaactaa	6120
accatgtgtt aagttgacac cactatgcgt aagtcttaag tgtacagacc taaaaacgaa	6180

-continued

tactaatact aactctagta gcggaagaat gattatggaa aaggagaga taaaaattg	6240
ttcattcaat attagtacat caattagggg taaggtacaa aaagaatacg cattttta	6300
taagtttagac ataataccaa tcgataacga tacaacatca tataagttaa ctatgttta	6360
tacttagcgtt attacacagg catgtcctaa ggtatcatc gaacctatac cgatacacta	6420
tttgtcacca gccggattcg caataactgaa atgtaataat aagacattt acggaaccgg	6480
accatgtact aacgtaagta cagtgcatt tacacacggg attagacccg tagtgagttac	6540
gaaattactg tttaaacggat cattagccga agaggaagtg gtaatttagat ccgttaattt	6600
taccgataac gctaaaacaa tttagtgca attgaatact agcgtagaga ttaattgtac	6660
tagacctaatt aataatacta gaaaacggat taggatacaa aggggaccag gtagggcatt	6720
cgttacaatc ggtaagatag ggaatatgag acaggccat tgtaatattt gtatggctaa	6780
atggaataat acacttaaac agatagctag taagctttaga gagcaattcg gtaataataa	6840
gacaattata ttcaaaacat ctatcgagggg ggttccagaa atcgtaacac atatgttta	6900
ttgcggaggc gaattttttt attgtatag tacacaattt ttcaatagta catggttta	6960
tagtacatgg tcaaccgaag gatctaataa tacagagggg tccgataccaa ttacactacc	7020
atgttaggatt aaacagataa ttaatatgtg gcaaaaaggta ggtaaggca tgtaatgcacc	7080
acctattatgc ggacaaatataa gatgtatcc gaatataacc ggattactgt taactagaga	7140
cggagggaat tcgaaataacg aatccgaaat ttttagacca gggggggggg atatgagaga	7200
caattggaga tccgaaattgt ataagtataa agtcgtttaaa atcgaaccat taggegtac	7260
acctacaaaaa gcgaaaagac gggtagtgca acgcgaaaaa agagccgtag ggataggcgc	7320
actattcttta gggtttttag ggcgcggccgg atcaactatg ggagcagctt gtatgacatt	7380
gacagtgcac gctagacaac tgtaagccg aatagtgcac caacagaata atctgttaag	7440
ggcaatcgaa gcacaacacg atctgttaca attgacagta tggggattt agcaatttgca	7500
ggctagaata ctgcgactcg aacggtatct gaaagaccaa caactgttag ggatatgggg	7560
atgttagcggt aagttgatat gtacaaccgc agtaccatgg aacgctatgtt ggtctataaa	7620
atcactcgaa cagatatggc atcatacaac atggatggaa tggatagag agattaacaa	7680
ttataactatgt ctgataact cactaatecg aagtcacaa accaaacagg aaaaaacgaa	7740
acaggaactg ttagatgtt acaaattggc tagcttatgg aattggttt atattactaa	7800
ttgggttatgg tatataaaac tggttattat gatagtaggg gggtagtgc gattgagaat	7860
cgtattcgca gtactatcaa tcgtaatag ggtagacag ggatatagtc cactatcatt	7920
ccaaacccac ctccccaccc cgaggggacc cgacaggccc gaaggaatag aagaagaagg	7980
tggagagaga gacagagaca gatccattcg attagtgaac ggatcttgg cacttatcg	8040
ggacgatctg cggagctgt gcctcttcag ctaccaccgc ttgagagact tactcttgat	8100
tgttaacgagg attgtggAAC ttctggacg caggggggtgg gaagccctca aatattggtg	8160
gaatctccta cagtatttggc gtcaggaact aaagaatagt gctgttagct tgctcaatgc	8220
cacagccata gcagtagctg aggggacaga tagggttata gaagtagtac aaggagcttgc	8280
tagagctatt cgccacatac ctagaagaat aagacaggcc ttggaaaggaa ttttgctata	8340
agatgggggggtaa tagtggctc aaatcttagcg taatcgatg gcctacagtt agggaaagaa	8400
tgcgtagacg cgaacctgca gccgatagggg taggcgcgc tagttagggat ctcgaaaaac	8460
acggagcaat tacttagtagt aatacagccg caactaacgc agcatgcgc tggttagagg	8520

-continued

cacaggaaga ggaagaggta gggttcccg taacaccaca ggtaccacctt agacctatga	8580
catataaagc cgcaagtgcgat ctatcacatt ttctgaaaga gaaagggggg ttagaggac	8640
tgatacatc acagagacga caagacatac tagacttatg gatatatcat acacaggat	8700
atttccaga ctagaaaaat tatacaccag gaccaggcg aagatatcca ttgacatcg	8760
gatggtgtt aaaactggta ccagtcgaac cagacaaaat cgaagaggct aacaaaggcg	8820
aaaatactag tctgttacat cccgttaaact tacacggat ggacgatcca gaaagagagg	8880
tactcgaatg gagattcgat agtagactcg cattccatca cgtagctagg gaattgcac	8940
cagagtattt taaaattgt tgacatcgag cttgctacaa gggacttcc gctgggact	9000
ttccagggag gcgtggcctg ggcgggactg gggagtgccg agccctcaga tcctgcata	9060
aagcagctgc ttttgcctg tactgggtct ctctggtag accagatctg agcctggag	9120
ctctctggct aactaggaa cccactgctt aagcctcaat aaagcttgcc ttgagtgtt	9180
c	9181

<210> SEQ ID NO 37
<211> LENGTH: 3309
<212> TYPE: DNA
<213> ORGANISM: Rotavirus sp.

<400> SEQUENCE: 37

ggctaaaaaaaa aatggcgcaa tcgatagtag tagatggta ttatgtgcg ttagcgtcgc	60
gttatttaaa atttgtatat gatttgaaa atgtaacgta taaaataat tatttgcga	120
cggataaatt taaaaagat atagaacaat attaaaatc gatacatgat ggtaaaaaaa	180
taacgcaatc gaaaatagat gaaaaagaaa aaatattatt agatcgtgta cggcggaaag	240
aacgttgttt aatatcgaaa ttatgttttgcgtatggta acatggtaat gtggaaaata	300
aatttagtaaa atatgggtta aaagatgcgt tatcgcatgc gcccggaaaa gatgcgaaac	360
cgtatggaaa taatataata acgtcgaaa tatttaaga aaaatcgaa tatacgata	420
tatatatggta tccgtcgata aatacgctgt gtcaatcgaa ttgtcaagcg atgatgttta	480
cgtatcgaa aatggaaatta aataatataa aaaatgcggc ggggttagaa aaattatcta	540
cgataatagc ggcgacgata aataaatatg gtatgcccg tcataatacg cggtatcggt	600
atgaatggga aacgatgaaa aataaaccgt atcatggatc ggcgtggata aattcgtcga	660
tagaaatgat agcgtgtgta gttagatcatc atacgtatata gatagcgcgt gaattaaatg	720
taaaatcggtt tacgaatcgat acgtcggttag cggaaattgtt atcgatcgccg atgacggat	780
taacggcgat gttaccgata cgtggtagt ttataacgac gggaaattta gaattagaat	840
attcgaataa atcgataatattat tatttaataat cggaaagaaat ggcggaaagat tttatgcaag	900
cgataaaaca attacgtgat gaaagggttag aatataacc ggattattat gaaaaatgg	960
ttaaaatcgcc ggatccgtta acgtttccga atatgcgtt aatataattcg ttttcgttcc	1020
atgttaggtt tcgtaaacaa gcggtatcg atgcggatata tgatcaaata acggtaacgt	1080
attcggataa tgtaaatatg gaaatgtata aagaatattc ggaacgtata gaaaatgaaa	1140
tatttacgtt attaaaagat aaaataatac atgaagataa acgttttagaa gaatatgaaat	1200
tatcgccgtt attatcgatc tcgtcgccgt cgaatggat attacgtgaa ataaattttg	1260
gtgggtcaaaa agtacgtcg acggaaaaaaa atatgcgtt aatagatgtt atatatacata	1320
aaaaatataac gacggatata cccgggttag atgcgcgtaa tccgataaccg ttaggtcg	1380
gtgtatgtacc gggtcgtcg acgcgtgcga tattttatattt accgtatcaa tattttatag	1440

-continued

cgcaacattc gtttgccgaa ataatgttaa attatgcgaa acgtgaacgt gaatattcg 1500
 aattttatc gcaagcgaat caagtattat cgtatggta tgtaacgcgt tatttagatt 1560
 cgaattcgat attatgttt acggatgtat cgcaatggga tgcgtcgcaa cataatacga 1620
 aagtattacg tcgttcgata atacgtgcga taaaacgttt aaaacaatta acgcataata 1680
 taaaataaca taaagcgata aatatataata tacaatcgca agaaaattta gaaaattcgt 1740
 atgttataat agataaaaaaa gcgatacaat atggtgcgcac ggcgtcggtt gaaaaacaaa 1800
 cggaaataat gaattcgata gcgatataag cgttaataca aacggtatta ggtaaattaa 1860
 tgcggattt tacgttgat gtaaaaatga tacgtgtaga tggtgatgat aattatgcga 1920
 tagtacgttt tccgatagcg ataacggaaa aattattatc ggaatttacg tcgaaattac 1980
 gttcgttata ttccggaaatg aatgtaaaag taaaacgtt aacgtcggtt acgggttgt 2040
 aaatagcgaa acgttatgtt gccccgtggta tggatatttt tgcgtcggtt gtaaaatata 2100
 tacatcatga aaaacgtaat caagattcg cgtatgatata ggcggcgacg ttatatgcga 2160
 attatatagt aaatgcgtt cgtggttaa cgtatgcgcg tacgtttata ttaacgaaaa 2220
 tatgtcaaat gacgtcgata aaaataacgg gtacgttacg ttatattccg atgaaatcga 2280
 tattagcgtt aaattccggcg tttaaagtat ttgtatgaagt agattatgtt ataaaattatc 2340
 cgatatcgaa ttatattata caattacaac gtaaattatc gtcgataaaaa gcgaaatcga 2400
 aaatagcgga taatatacgaa aatcgccgc aatttaaattc gtatgttagaa ttatataata 2460
 aatcgtaac gacggatgaa aatccgatag tatacgatgg tatacgatgg acggaaaaag 2520
 cggaaattaaa ttgcgtatgcg ccgatagcgat tagaaaaacg tgcgtatcaa ttatcgat 2580
 tggatcgatgtt ttacaaaat ccgacgacgt ttaaattcgaa aacggatgtt acgatataatg 2640
 atgttattata ttatattatcg gggtttataa aaatagattt gtcgacggta ttaccgaaag 2700
 aagaaaaataa tacgtatccg ttattaccgg cgataataaa acgtacgtt tcgtatattt 2760
 gtttacgtac gcatgattat gatataaaag gttcgtcgat gacggatcg aaaataataa 2820
 aacaatattc ggtatatacg ccgggtatag aagaattata tggatgttata aataaattcg 2880
 tagatacgat acgtggatattt ttgcgtatgtt ttaatgttacc gaaageggat gtagatacg 2940
 atatatcgac gcaaatgtat aaatcgatc gttttaaat attacaacgcg tataatata 3000
 atttattatc ggtaaattat ggtatgtatc aattatgttata tttaattatcg ggcgttttt 3060
 ttgcgtatgtt aatacatatcg ccgatggcgaa aacgcggac ggcgttattt atgatagatt 3120
 tagcgttacg tttaaaaata ataaatcatt gtatagaaaa aggtgaaata ataaacggat 3180
 cggatcgatgc gaataaaaacg gattatattaa aattatggcg tgcgtatgg aatgtaaaa 3240
 cgtatgttccg gccgtatccg aaaaattcga tggatgttata gaaagagaag tggattgcatt 3300
 attgtggct 3309

<210> SEQ_ID NO 38
 <211> LENGTH: 3309
 <212> TYPE: DNA
 <213> ORGANISM: Unknown
 <220> FEATURE:
 <223> OTHER INFORMATION: synthetic

<400> SEQUENCE: 38

ggctaaaaaa aatggcacaa tcaatcgtag ttgacggaga ttatgtatgcg ttagctatgtt 60
 gatatctaaa attcgatatac gatttcgaaa acgtacttata tcagaataat tatttcgcaa 120

-continued

-continued

ctaaattgaa ttcatacgca ccaatcgcat tagaaaaacg tagagatcaa ttttcaatta	2580
tggtatcatt ttacagaat cctactactt ttaaatccga aacagttagt acaattaacg	2640
atgtattgtt ttttatatca ggatttataa aaatcgattc tagtacagtg ttacctaaag	2700
aggaaaataa tactatgcc aattaccag ctataattaa acgtacacta tcataattcg	2760
gattgcgtac acatgattat gatattaagg gatcatctag tacagtatcg aaaattataa	2820
aacagtattc agtataataca ccaggaatcg aagagttata cgaaatcggtt aataaatcg	2880
tagatacaat tagggatat tttgttagtt ttaacgtacc taaagctgtat gtagatacat	2940
atatatcgac acaaattgtat aaacatgata gatttaaat actgcaagca tatataatata	3000
atctgttatac agttaattat ggtatgtatc aattagtcga tctgaattca gctagattt	3060
ttgatcatgt aatacatatac cctatggcta aaacacccatc agcagtttt atgatagatt	3120
tagcgttaag attgaaaatt ataaatcatt gtatcgaaaa aggccaaatt ataacagtt	3180
ggttacacgc taataaaact gattatctt agttatggcg tatgttatgg aacgttaaga	3240
ctatgaatag tccatattct aaaaattcaa tgtttgacga ataagagaag tggattgcat	3300
attgtggct	3309

<210> SEQ ID NO 39
<211> LENGTH: 2736
<212> TYPE: DNA
<213> ORGANISM: Rotavirus sp.

<400> SEQUENCE: 39

ggcttaaaaa gatcagttga ggacaaatcg ttcaagatga ttcgttgtaa tcgttgtcgt	60
aataatcaac aaaaaatatac agaaaaagaa aaacaattatc aaacgataat aaataaagaa	120
gtaaaaagaaa ataaagattc gatgaaagaa gataaattatc tagtaacggaa agaatcgat	180
gggtgtatgtt caacggcgaa agaacaatcg aataatataa atttacaaaa aatgtat	240
gtaaaaagaag taatgaatat acaaaatcaa acgttaatataa cggtagtaac ggaaaataaa	300
gtagaaaatg aagaaaatagt aaaaaatatac ataccgtcgta ataaatcgaa ttccgttata	360
gtaaaaaaaaat taacggaaat acaagaatcg tcggcgaaaa cgtataatatac gttatccgt	420
ttatccgttccgatc cggtaaaatc gtatccatatac gatataatcg gtgaaaaaaa attatcgacg	480
cgttgttattt ggaaattttt aaaaatgtat ttaccggcggtt gttttatcc ggtacgttac	540
ttttttttat cgttataatccatatac aatgtatataa gaagaaatgc cggattatatac aatgttacgt	600
gatatggcggtt tagataatccatatac gtattccggcg gaagcggttataa aatgttacgt	660
aaagaaaatataatc taatagaatataatcgat caaatgcggg aaggttatatac acgtcgat	720
atgtcgaaat tacgttataa aatatcggtt gaaacgaaataa cggcgaaataa tccggcgata	780
ttacatccggat tagataatgcatatac aatgttacgtt aacatcaatt aatacaaccg	840
ttaacgcgcgat gtaatataatcg ggaatataataa cccgacgcataatataatcgat tccgttataat	900
gtatccatatac tagatgcggc gtttttaacg aatttcgttccgtt ttgttacccgtt gtatccatatac	960
caagatgttataa taggttttataa tttttttttt gtttttataa aatgttacgtt aacatcaatt aatacaaccg	1020
gattatgtat cggcggttccgtt tttttttttt gtttttataa aatgttacgtt aacatcaatt aatacaaccg	1080
caaaaatggcggtt cggcggttccgtt tttttttttt gtttttataa aatgttacgtt aacatcaatt aatacaaccg	1140
caatccatatac taggttttataa tttttttttt gtttttataa aatgttacgtt aacatcaatt aatacaaccg	1200
tccgttataatcgat cggcggttccgtt tttttttttt gtttttataa aatgttacgtt aacatcaatt aatacaaccg	1260

225

226

-continued

```
<210> SEQ ID NO 40
<211> LENGTH: 2736
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: synthetic
```

<400> SEQUENCE: 40

ggcttaaaaa gatcagttga ggacaaatcg ttcaagatga tatctagaaa tagacgtaga 60
aataatcgc aaaaaaatat cgaaaaagaa aaacagttag agacaattat taataaagag 120
gttaaagaga ataaggattc aatgaaagag gataagttag tggtaaccga agaatctaac 180
ggagacgtta caaccgctaa agagcaatct aataatatta atctgcaaaa aaacgattta 240
gttaaagagg ttatgaatat acagaatcag acacttaata cagtagttac cgaaaataag 300
gttgagatag aagagatagt taaaaaatat ataccgtcat ataatactga ttcattaatc 360
gttaaaaaat taaccgaaat acaggaatct agtgctaaaa ctatataaac actatttaga 420
ttgtttcacac cagttaaatc atacttatac gatattaatg gcgaaaaaaaaa actatctact 480
agatggtatt ggaaactgtt aaaagacgt ctaccagcag gtgattactc agttagacaa 540
ttttttctat cattataactt aaacgttata gagggaaatgc cagattatata tatgctttaga 600

-continued

gatatggcag ttgataatcc atactcagct gaagcaggta agatagttga cggtaaatct	660
aaagagatat taatcgatt gtatcaggat caaatgcacg agggatataat tagacggtat	720
atgtcagaac ttagacataa aatttcaggc gaaactaataa ccgctaaataa tccagctata	780
ctgcatccag tagataacga attgaatcaa tattttctcg aacatcaatt gatacaacca	840
ttaactacta gaaatatacg cgaattgata ccgacacaat tgtatcatga tcctaattac	900
gttttaata tagacgctgc atttttaact aattctagat tcgttaccacc atacttaacg	960
caagatagga tagggttaca tgacggattc gaatcaattt gggattctaa aacacatgct	1020
gattacgtat cagctagacg atttataccc gattnaaccg aatttagttga cgctgaaaaa	1080
cagattaagg aaatggctgc acatttacaa ctgcgaageta ttacggtaca gggtgaatca	1140
caatttttag ccggatttag cgccggcgcga gctaaccgat catttaattt tattatagga	1200
tcaatgttat caactagaac tatagcagtc gaattttatta catctaatta tatgtcatta	1260
gttagttgtat tttttttttt gacaattttt ccttagcgaaa ttttttttttt cgaatcattta	1320
gtcgctatgc aattagctat tattaataca cttatatacc cagcattagg gttacgcgaa	1380
atgcattatc aggccaggta gggttacaca ccatttcgaat tagccgaaat gcaaggttgcg	1440
aatagatcaa ttagacaatg gttacatcat tgtaatacat tacaattcgg taggcaattt	1500
accgaaggta ttatacatct tagatttact aacgatatta tgacaggtag gatagttaat	1560
ctatttcga ctatgttagt tgctttatca tcacaaccat tcgctacata tccatttagat	1620
tataaaagat cagtgcacg agcattacaa ttgttatcta atagaaccgc tcaaatacgct	1680
gatttaacta gactgatagt gtataattt actacattat ccgcattgcgcat agttatgaat	1740
atgcatttag tgggttact tacagtcgaa cgtatacagg ctacatcattt aacatcattt	1800
atgatgttaa tttctataaa aacagttata cccgaaccat catcattttt ttcatatttt	1860
tcatctata ttaattttt aactaattt aacgaacaga tagataacgt agttgcggaa	1920
attatggctg catatagattt gaatctatataa caacaaaaaa tgtaatgtt agtgcactaga	1980
ttcgttaagta agttatataat attcgatcga cctaaaattt caccagatca aatgtataga	2040
cttagaaata gacttagaaa tataccagtc gaaagacgta gagcagacgt atttagaatt	2100
attatgaata atagggattt aatcgaaaaa actagcggaa gaatttgta gggagttta	2160
ctatcatata cacctatgcc attaacatac gttgaagacg taggattaac taacgttaatt	2220
aacgatacta attcggttca aattttaat atcgaagaaa tcgaaaaaac aggtgattac	2280
tcaatgttata ctaatgcact attaaggat acaccttata tacttaaggg agctatacc	2340
tacgttacga attcgctcgt gatagacgta ctatctaagg tagatactac agtattcgct	2400
agtatagtta aggataggta tataatctaag ttaaaaccaaa ttaaaatttata aattaattct	2460
gattctagcg aatattttt agtgcataat aataaaatggta cacctactac tactactcga	2520
gtgtataaag ctagatcaca acaattcgat atacaacata gcgttatctat gctcgaatct	2580
aatctatttt tcgttagtgta taacgatcta tttaaatata ttaaaactac tacagtgtta	2640
ccaattaacg ctgttatcata cgtggagot agaattatgc aagaaacata aatgattgtta	2700
tagtatcatc ttgtgacgac ctcaaactttt gtggct	2736

<210> SEQ ID NO 41

<211> LENGTH: 29751

<212> TYPE: DNA

<213> ORGANISM: Human coronavirus

-continued

<400> SEQUENCE: 41

-continued

-continued

tcggagaaac attttgtaga aacggtatcg tttagcgggtt cgtatcgta ttggtcgtat 4740
tcgggtaac gtacggatt aggtgtagaa tttttaaac gttggtataa aatagtatat 4800
catacgtag aatcgccgt agaatttcat ttagatggtg aagtatttac gtttagataaa 4860
ttaaaatcg tattatcggt acgtgaagta aaaacgataa aagtatttac gacggtagat 4920
aatacgaatt tacatacgca attagtagat atgtcgatga cgtatggtca acaatttgg 4980
ccgacgtatt tagatggtgc ggatgtaaacg aaaataaaac cgcatgtaaa tcatagaaagg 5040
aaaacgttt ttgttattacc gtcggatgt acgttacgtt cggaaagcggt tgaatattat 5100
catacgtag atgaatcggt ttttaggtcgt tataatgtcg cgttaaatca tacgaaaaaa 5160
tggaaatttc cgcaagtagg tggtttaacg tcgataaaaat gggcggataa taattgttat 5220
ttatcgctgg tattattagc gttacaacaa ttagaagtaa aatthaatgc gcccggctta 5280
caagaagcg attatcgtagc gcgtgcgggt gatgcggcga attttgcgtc gttaatattat 5340
gcgtattcga ataaaacggt aggtgaattt ggtgtatgtc gtgaaacgat gacgcattta 5400
ttacaacatcg cgaattttaga atcggcgaaa cgtgtattaa atgttagtata taaacattgt 5460
ggtcaaaaaa cgacgcacgtt aacgggtgtaa gaagcggtaa tggatatggg tacgttatcg 5520
tatgataatt taaaaacggg tggatcgata ccgtgtgtat gttgtcgta tgccgacgcaa 5580
tatttagtac aacaagaatc gtcgtttgtaa atgtatgtcg ccgcggccgc ggaatataaa 5640
ttacaacaag gtacgtttt atgtgcgaat gaatatacg gtaattatca atgtggtcat 5700
tatacgata taacggcgaa agaaaacgtta tatacgatag atggtgcgcata tttaaacgaaa 5760
atgtcgaaat ataaagggtcc ggttaacggat gtattttata aagaacgtc gtatacgacg 5820
acgataaaaac cggtatcgta taaatttagat ggtgtaaacgt atacggaaat agaaccgaaa 5880
ttagatgggtt attataaaaaa agataatcg tattatacg aacaaccgat agatttagta 5940
ccgcgcacac cggttaccgaa tgccgtgtt gataatttta aattaacgtt ttcgaatacg 6000
aaatttgcgg atgatttaaa tcaaatgcg ggtttacgaa acccggegtc gcgtgaattta 6060
tcggtaacgt tttttccggtt tttaaatggt gatgttagtag cgatagattta tcgtcattat 6120
tcggcgctgt taaaaaaagg tgccaaatata ttacataaac cgatagatgt gcatataat 6180
caagcgacac cgaaaacgac gtttaaccgaa aatacgtgtt gttacgtgtt tttatggtcg 6240
acgaaaccgg tagatacgtc gaattcggtt gaagtattag cggttagaaaga tacgcacgg 6300
atggataatt tagcgtgtga atcgcacaa ccgcacgtcg aagaagtagt agaaaatccg 6360
acgataacaaa aagaagtaat agaatgtgtat gtaaaaacgaa cggaaagtagt aggttaatgt 6420
atattaaaaac cgtcggatgtaa aggtgtaaaa gtaacgcacaa aatttaggtca tgaagattta 6480
atggccggcgtt atgttagaaaaa tacgtcgata acgataaaaaa aaccgcacaa attatcgta 6540
cggttaggtt taaaaacgat agcgcacgtt ggtatagcggtt cgttaatatttca ggttccgtgg 6600
tcgaaaatata tagcgtatgtt aaaaaccgtt ttaggtcaag cggcgataac gacgtcgat 6660
tgtgcgaaac gtttagcgca acgtgtatttta aataattata tggcgatgtt attacgtta 6720
ttatttcaat tatgtcgatgtt tacgaaatcg acgaaattcg gtaacgtgtc gtcgttaccg 6780
acgacgtatcg cggaaaatccg ggtttaccgtt gtagcggaaat tatgttttaga tggcggtata 6840
aattatgttaa aatcgccgaa attttcgaaa ttatttacgaa tagcgtatgtt gttattattat 6900
ttatcgatat gtttaggttc gttaaatatgt gtaacggcggtt cgtttgggtt attattatcg 6960
aattttgggtt cggccgtcgta ttgtatgggtt gtagcgtgtat gtaatttaaa ttcgtcgat 7020
gtaaacgcacac tggatgtttt gtaaggttgcg tttccgtgtt cgtatgtttt atcgggtttt 7080

-continued

gattcgtag attcgtagatcc ggccgttagaa acgataacaag taacgatatac gtcgtataaa	7140
ttagatttaa cgatattagg tttagcggcg gaatgggtat tagcgtatata gttatttacg	7200
aaatttttt atttattagg tttatcggcg ataatgcaga tattttttgg ttatttgcg	7260
tcgcattta tatcgaattc gtggtaatg tggttataa tatcgtatgt acaaatggcg	7320
ccggtatcgg cgatggtacg tatgtatata tttttgcgt cgttttatta tatatggaaa	7380
tcgtatgtac atataatgga tgggtgtacg tcgtcgacgt gtatgtatgt ttataaacgt	7440
aatcgtgcga cgcgtgtaga atgtacgacg atagtaatgt gtatgaaacg ttctgtttat	7500
gtatatgcga atgggtgtcg tgggtttgtt aaaacgcata attggaatgtt tttaaatgt	7560
gatacgtttt gtacgggttc gacgtttata tcggatgaag tagcgcgtga tttatcgta	7620
caatttaaac gtccgcataaa tccgcacggat caatcgtcgat atatagttaga ttccggtagcg	7680
gtaaaaaatg gtgcgttaca tttatatttt gataaagcgg gtcaaaaaac gtatgaaacgt	7740
catccgttat cgcattttgtt aaatttagat aatttacgtg cgaataatac gaaagggtcg	7800
ttaccgataa atgtatagt atttgatgtt aaatcgaaat gtatgatgc ggcgtcgaaa	7860
tcggcgtcg tatatttttc gcaattaatgt tgcgtacccgat tattttattt agatcaacgc	7920
tttagtacgg atgttaggtt ttcgcacggaa gtatcgtaa aatgtttgt tgcgtatgt	7980
gatacgtttt cggcgacgtt ttcgggtacccg atggaaaaat taaaagcgtt agtagcgacg	8040
gcccattcgg aattagcgaa aggtgtacgt ttagatgggtt tattatcgac gtttgcgtat	8100
ggggcgcgtc aagggtgtgtt agatcgatcc gtagatacgaa aagatgtat agaatgttta	8160
aaattatcgc atcattcgaa tttagaagta acgggtgattt cgtgtataaa ttttatgttta	8220
acgtataataa aagttagaaaa tatgacgcgc cgtgatttagt gtgcgtgtat agattgtat	8280
gcgcgtcata taaatgcgc aatcgacggaa tcgcataatgt tatcgtaat atggatgtt	8340
aaagattata tgcgttatac ggaacaattt cgtaaacaaa tacgttgcgc ggcggaaaaaa	8400
aataatatac cgtttcgtttt aacgtgtcgcc acgacgcgtc aagtagttaaa tgcgtatgtt	8460
acgaaaaat cgtttaaagg tggtaaaataa gtatcgacgt gttttaaattt aatgtttttttt	8520
gcgcgttata tgcgtgtattt agcggcgatcc gtatgttataa tagtaatgcc ggtacatcg	8580
ttatcgatac atgtatggtaa tacgtatgtt aatatagtt aatggatgtt acaagatgggt	8640
gtacgcgtg atataatatc gacggatgtatgtt tgcgtatgtt aatggatgtt acaagatgggt	8700
gcgtgggtt cgcacgtgg tgggtcgat taaaatgtt aatcgatgtcc ggtatgtacgt	8760
gcgataataa cgcgtgaaat aggtttataa gtaccgggtt taccgggtac ggtattacgt	8820
gcgataataatgtt gttttttttaa acatgtttaa ccgcgtgtat tttcgccgtt aggtatata	8880
tgttatacgcc cgtcgaaattt aatagaatataa tgcgtatgtt tgcgtatgtt acaagatgggt	8940
ggggcggaaat gtacgtatattt taaagatgcg atgggtaaac cgggtacccgtt ttgttgcgtatgtt	9000
acgaattttat tagaagggttc gatatcgatcc tcggatatttac gtcggatatac gcgatgtatgtt	9060
ttaatggatgtt tgcgtatgtt acaatttccg aatacgatcc tgcgtatgtt acaagatgggt	9120
gtacgcgttata tgcgtatgtt aatggatgtt acaatttccg aatacgatcc tgcgtatgtt acaagatgggt	9180
atatgtttat cgcgtatgtt tgcgtatgtt aatggatgtt acaatttccg aatacgatcc tgcgtatgtt acaagatgggt	9240
gggtgtatgtt tgcgtatgtt aatggatgtt acaatttccg aatacgatcc tgcgtatgtt acaagatgggt	9300
caaccggtagt gtcgtatgtt aatggatgtt acaatttccg aatacgatcc tgcgtatgtt acaagatgggt	9360
ttagtaacgt gtcgtatgtt aatggatgtt acaatttccg aatacgatcc tgcgtatgtt acaagatgggt	9420

-continued

-continued

ttacaacaat tacgtgtaga atcgtcgaa aaattatggg cgcaatgtgt acaattacat 11880
aatgatatat tattagcgaa agatacgacg gaagcggtt aaaaaatgggt atcggttatta 11940
tcgggttattat tatcgatgca aggtgcggta gatataaatac gtttatgtga agaaaatgtta 12000
gataatcgta cgacgttaca agcgatagcg tcggaatttt cgtcggttacc gtcgtatgct 12060
gegtatgcga cggcgcaaga agcgtatgaa caagcgtag cgaatgggtga ttccggaaagta 12120
gtattaaaaa aattaaaaaa atcggttaat gtacgcaaat cggaaatttga tcgtgtatgct 12180
gegatgcac acgtttataga aaaaatggcg gatcaagega tgacgcaaat gtataaaca 12240
gcccgttcgg aagataaaacg tgccaaagta acgtcgccga tgcaaaacgat gttatttacg 12300
atgttacgta aatttagataa tgatcggtta aataatataa taaataatgc gcgtgtatgg 12360
tgtgttacgt taaatataat accgttaacg acggcgccga aattaatgggt agtagtaccg 12420
gattatggta cgtataaaaa tacgtgtat ggtatatacg ttacgtatgc gtcggcggtta 12480
tggaaatac aacaaggtagt agatcgccgt tcgaaaatag tacaattatc ggaaataaaat 12540
atggataatt cggccgaattt acgtcgcccg ttaatagtaa cggcggttacg tgcaatttcg 12600
gccccggaaat tacaatataa tgaattatcg ccggtagcgta acgtcaaat gtcgtgtcg 12660
gccccgtacg cgcggaaacggc gtgtacggat gataatcgct tagcggttata taataattcg 12720
aaagggtggtc gttttgtatt acgttattat tcggatcatc aagattaaa atggcgccgt 12780
tttccggaaat cggatggta cgggtacgata tatacgaaat tagaaccggcc gtgtcggttt 12840
gttacggata cggccgaaagg tccggaaatgta aaatattttt attttataaa aggtttaat 12900
aattttaaatc gtggatgtt attaggttcg ttacgtcgccga cggtagcggtt acaaggccgt 12960
aatgcgacgg aagtaccggc gaattcgacg gtattatcg tttgtcggtt tgccggtagat 13020
ccggcgaaag cgtataaaga ttatggatcg tcgggtggtc aaccgataac gaattgtgt 13080
aaaatgttat gtacgcatac gggtagggta caagcgataa cggtaacgccc ggaagcgaat 13140
atggatcaag aatcggttgg tggtgctcg tggatgttattt atgtcggttgc tcatatagat 13200
catccgaatc cgaaagggtt tggatgttta aaaggtaat atgtacaat accgacgacg 13260
tgtgcgaatg atccggtagg ttttacgtta cgtataacgg tatgtacggt atgtggatgt 13320
tggaaagggtt atgggtgttc gtgtatcaa ttacgtgaac cgttaatgc atccggcgat 13380
cgctcgacgt ttttaatgg gtttgcggta taagtgcggc cggttaacg ccgtgtggta 13440
ccgggtacgtc gacggatgtta gtatcgatcg cgttgcata atataatgaa aaagtacgccc 13500
gttttgcgaa atttttaaaa acgaattgtt gtcgtttca agaaaaagat gaagaaggta 13560
atttatttgcgaa atttttaaaa acgaattgtt gtcgtttca agaaaaagat gaagaaggta 13620
aaacgtatata taatggatcg aatggatgtc cggcggtacg ggtacatgtat ttttttaat 13680
ttcgtgtatcg tggatgtatcg gtaccgcata tatacgatcg acgtttaacg aatatacg 13740
tggcggttacgtatcg ttacgtcatt ttgtatgcagg taattgtatcg acgtttaacg 13800
aaatatttgcgaa attttgcgaa attttgcgaa acgtttaacg aatatacg aatatacg 13860
attttgcgaa attttgcgaa attttgcgaa acgtttaacg aatatacg aatatacg 13920
aaatcggttata aaaaacggta caatggatcg atgcgtatcg tgatgcgggt atgttaggt 13980
tattaaatcgta agataatcaa gatggatcg tattttgtatcg gttttgtatcg 14040
aagtacgccc ggggttgcgtt gtaccgtatcg tagattcgta ttattcggtta ttaatgcgg 14100
tattaaatcgta aacgcgttgcgtt ttacggccgg aatcggtatcg ggatgcggat ttacggaaac 14160

-continued

cgttaataaa atgggattta taaaatatg atttacgga agaacgttta tgtttattt 14220
 atcgttattt taaatatgg gatcaaacgt atcatccgaa ttgtataaat tgtttagatg 14280
 atcgttgtat attacattgt gcgaatttta atgtattatt ttcgacggta tttccgcga 14340
 cgtcgttgg tcgcgttagta cgtaaaatat ttgtagatgg tgtaccgtt gtatgtatcga 14400
 cgggttatca tttcgtgaa ttaggtgttag tacataatca agatgtaaat ttacattcgt 14460
 cgcgttatac gttaaagaa ttattgtat atgcggggaa tccggcgatg catgeggcgt 14520
 cgggttaattt attattagat aaacgtacga cgtgttttc ggtageggcg ttaacgaata 14580
 atgttagcgtt tcaaaccgtt aaaccgggtt attttaataa agattttat gatttgccg 14640
 tatacgaaagg ttttttaaa gaaggttcgtt cggtagaatt aaaacatttt tttttgcgc 14700
 aagatggtaa tgcggcgata tcggattatg attattatcg ttataattt ccgacgatgt 14760
 gtgatatacg tcaattatta tttgttagtag aagttagtataaataattt gattgtttag 14820
 atggtggttg tataatgcg aatcaagtaa tagtaataaa tttagataaa tcggcgggtt 14880
 ttccgtttaa taaatggggtaa aaagcgcgtt tatattatga ttgcgtatcg tatgaagatc 14940
 aagatgcgtt atttgcgtat acgaaacgtt atgtaataacc gacgataacg caaatgaatt 15000
 taaaatatgc gatatccggc aaaaatcgtt cgcgtacggt agcgggtgtt tcgatatgtt 15060
 cgacgatgac gaatcgtcaa tttcatcaaa aattattaaa atcgatagcg ggcacgcgt 15120
 gtgcgacggt agtaataggt acgtcgaaat tttatgggtt ttggcataat atgttaaaaa 15180
 cggtatatttcc ggtatgttagaa acgcccattt taatgggtt ggattatccg aatgtgatc 15240
 gtgcgatgcc gaatatgtt cgtataatgg cgtcggtatg attagcgcgtt aaacataata 15300
 cgtgttgcataa ttatcgcat cgtttttatc gtttagcgtt tgaatgtgcg caagtattat 15360
 cggaaatggt aatgtgttggt gggtcggtt atgtaaaacc ggggtggtacg tcgtcgggt 15420
 atgcgacgac ggcgtatgcg aattcggat ttaatataatg tcaagcgtt acggcgaatg 15480
 taaatgcgtt attatcgacg gatggtaataa aaatagcggta taaatatgtt cgtatattac 15540
 aacatcgaaa atatgaatgt ttatatcgta atcgatgtt agatcatgtt tttgttagatg 15600
 aattttatgc gtatattacgtt aaacatttt cgtatgtt attatcgatgat gatgcggtag 15660
 tatgttataa ttcaattat gcccccaag gtttagtgcgtt gtcgataaaa aattttaaag 15720
 cggattataa ttatcaaaaat aatgtattta tgcggaaacg gaaatgttgg acggaaacgg 15780
 atttaacgaa aggtccgcattt gatgttgcgtt cgcacatcgtt gatgttgcgtt aaacaagggt 15840
 atgattatgtt atatttaccg tattccggatc cgtcgctat attaggtgcgtt ggttgggg 15900
 tagatgtatgat agtaaaaaacg gatgggtacgtt taatgtatgaa acgttttgcgtt tcgttgcgtt 15960
 tagatgcgtt tccgttacgc aaacatccgtt atcaagaataa tgcggatgtt tttcatttt 16020
 atttacaataa tatacgtaaa ttacatgtatg aatatacggg tcatatgtt gatgttgcgtt 16080
 cggtaatgtt aacgaatgtt aatacgtcgtt gttattggaa accggaaattt tatgaagcgtt 16140
 tgcgtatcgcc gcatacggta ttacaaggcg taggtgcgtt tgcgttgcgtt aattcgcaaa 16200
 cgtcggttacg ttgtgggtgcgtt tgcgtatcgatc gtcgttgcgtt atgttgcgtt gtttttttt 16260
 atcgtatgtt atcgacgtcg cataaattttt gatgttgcgtt aaatccgtat gatgttgcgtt 16320
 cggccgggttgcgtt gatgttgcgtt gatgttgcgtt aattatattt aggtgggtatg tcgttgcgtt 16380
 gtaaaatgcgtt taaaccggccg atatcgatcgtt cgttgcgtt gatgttgcgtt gtttttttt 16440
 tataatgggttgcgtt gatgttgcgtt gatgttgcgtt aatgttgcgtt atgttgcgtt 16500
 gtgattggac gatgttgcgtt gatgttgcgtt gatgttgcgtt gatgttgcgtt gtttttttt 16560

-continued

tatttgcggc ggaaacgtta aaagcgacgg aagaaacgtt taaattatcg tatggtag 16620
cgacggtagc tgaagtatta tcggatcgta aattacattt atcgtggaa gtaggtaaac 16680
cgcgccgccc gttaaatcgt aattatgtat ttacgggtta tgcgtgtaaacg aaaaattcg 16740
aagtacaat aggtgaatat acgtttgaaa aagggtgatta tggtgatgcg gtagtatatc 16800
gtggtagc gacgtataaa tttaatgttag gtgattattt tgcgttacg tcgcatacg 16860
taatgcgtt atcggcccg acgttagtac cgcaagaaca ttatgtacgt ataacgggtt 16920
tatatccgac gttaaatata tcggatgaat ttgcgtcgaa tgcgttacg tatcaaaaag 16980
taggtatgca aaaatattcg acgttacaag gtccgcggg tacggtaaa tcgcattttg 17040
cgataggtt agcgttatata tgcgttacg cgcgtatagt atatacggcg tggttcgc 17100
cgccggtaga tgcgttacg gaaaaagctg taaaatattt accgatagat aatgttcgc 17160
gtataatacc ggcgcgtgcg cgtgttagat gtttgataa atttaagta aattgcacgt 17220
tagaacaata tgcgttacg acgttacaatg cggttacggg aacgacggcg gatatagtag 17280
tatttgcgtt aatatcgatg gcgcacgattt atgatttttgcgttacg gatatagtag 17340
gtgcgaaaca ttatgtatata tgcgttacg cggcgcattt accggcccg cgtacgtt 17400
taacgaaagg tgcgttagaa ccggatattt tgcgttacg tgcgttacg gatatagtag 17460
taggtccggat tgcgttacg gatgttgcg tgcgttacg ggcggaaata gatgtacgg 17520
tatcgccgtt agtataatgtat aataaattaa aagcgatata agataatcg ggcgttacg 17580
ttaaaatgtt ttataaagtgt aatgttacg atgtatgttgcgttacg gatatagtag 17640
aaatagggtt agtacgttacg ttttacggc gtaatccggc gtgggttacg ggcgttacg 17700
tatcgccgtt aatgttacg aatgttacg cgttacg gatgttacg gatatagtag 17760
cggttagattt tgcgttacg tgcgttacg gatgttacg gatatagtag 17820
cgccgcattt tgcgttacg aatgttacg gatgttacg gatatagtag 17880
tattatgtat aatgttacg gatgttacg gatgttacg gatatagtag 17940
taccgcgtcg tgcgttacg acgttacaag cggaaatgttacg gatgttacg gatatagtag 18000
gttcgaaaat aataacgggttacg tgcgttacg gatgttacg gatatagtag 18060
taaaattaa aacggaaatgttacg tgcgttacg gatgttacg gatatagtag 18120
atcggttttacg tgcgttacg gatgttacg gatgttacg gatatagtag 18180
atgttacg tgcgttacg gatgttacg gatgttacg gatatagtag 18240
tagaagggttacg tgcgttacg gatgttacg gatgttacg gatatagtag 18300
tttcgcgttacg tgcgttacg gatgttacg gatgttacg gatatagtag 18360
cggaattttac gatgttacg gatgttacg gatgttacg gatatagtag 18420
cggttacg tgcgttacg gatgttacg gatgttacg gatatagtag 18480
cggttacg gatgttacg gatgttacg gatgttacg gatatagtag 18540
aattaaacgttacg tgcgttacg gatgttacg gatgttacg gatatagtag 18600
ataaaacgttacg tgcgttacg gatgttacg gatgttacg gatgttacg 18660
taggttttacg tgcgttacg gatgttacg gatgttacg gatgttacg 18720
gtatattacat tgcgttacg gatgttacg gatgttacg gatgttacg 18780
cggttacg gatgttacg gatgttacg gatgttacg gatgttacg 18840
atgggtcggttacg tgcgttacg gatgttacg gatgttacg gatgttacg 18900

-continued

agaatcacaaca tatggtagta aaatccggcg tattagcgga taaattcccg gtattacatg 18960
atataggtaa tccgaaagcg ataaaatgtg taccgcaagc ggaagtagaa tgaaaaatttt 19020
atgatgcgca accgtgttcg gataaagcg ataaaataga agaatttattt tattcgtatg 19080
cgacgcacca tgataaattt acggatggtg tatgtttattt ttgaaattgt aatgtagatc 19140
gttatccggc gaatgcgata gtagtcgtt ttgatacgcg tgtattatcg aatttaaattt 19200
taccgggtt tgatgggtcg tgcgttatgt taaaataaca tgcgtttcat acgcggcg 19260
ttgataaattt ggcgtttacg aatttaaacc aattaccgtt tttttattat tcggattcgc 19320
cggtgtgaatc gcatggtaaa caagtagtat cggatataga ttatgtaccg ttaaaatcg 19380
cgacgtgtat aacgcgttgt aatttaggtg gtgcggatgc tgcgtcatcat gcgaatgaat 19440
atcgtcaata tttagatgcg tataatatga tgatatccgc gggttttcg ttatggatata 19500
ataaacaattt tgatacgtat aatttatggg atacgtttacg cggttacaa tcgttagaaa 19560
atgtagcgta taatgtagta aataaaggctt attttgcgtt tcatgcgggt gaagcgccgg 19620
tatcgataat aaataatgcg gtatatacga aagtagatgg tatagatgt aaaaatttt 19680
aaaataaacc gacgttaccc gtaaatgtcg cgttgcattt atggcgaaa cgtatataaa 19740
aaccgggttacc ggaaataaaaa atattaaata atttaggtgt agatatacg gcgaatacgg 19800
taatatggg ttataaacgtt gaagcgccgg cgcgtatcgc gacgataggt gtatgtacga 19860
tgacggatattt agcgaaaaaa ccgacggaaat cggcggttgc gtcgttaacg gtattatgg 19920
atggctgtgtt agaaggctt gtagattttt ttcgtatgc gcgtatgtt gtattataaa 19980
cggaagggttcc ggtaaaagggt ttaacgcgtt cggaaagggttcc ggccaaagcg tcggtaatgt 20040
gtgtacgtt aatagggttca tcggtaaaaaa cgcaattttt ttatttttt aaagtagatg 20100
gtataataca acaattaccg gaaacgtattt ttacgcatacgc gctgtatatttta gaagatttt 20160
aaccgcgttcc gcaaatggaa acggatttt tagaatttgc gatggatgaa ttatataac 20220
gttataaattt agaagggtt gctgttgcattt atatagtata tcggatattt tcgcgttgc 20280
aatttagtggg ttacatttta atgatagggtt tagcggaaaccc ttgcgtatgc tcgcgtttaa 20340
aatttagaaga ttatataccg atggatttgcg cggtaaaaaaa ttatttttt acggatgcgc 20400
aaacgggttcc gtcgaaatgtt gtagttccgta taatagattt attattagat gattttgcgt 20460
aaataataaa atcgcgtatgc ttatcggtt tattcgaaatgtt agtaaaatgtt acgtatgtt 20520
atgcggaaat atcggtttatg ttatgggttta aagatggctt tgcgttgcgc ttatccgc 20580
aattacaacgc gtcgcgttcc tggcaaccgg gttacgttgc gcccgttataaataatgc 20640
aacgtatgtt attagaaaaaa tgcgttgcgtt aaaaattatgg tggaaatgcg gtaataccga 20700
aagggtataat gatgtatgtt gcaatgttgcg tcaatatttta aatacgatgtt 20760
cggttagcggtt accgtataat atgcgtgttac tacatgggttgc gtcgggttgc gataaagggt 20820
tagcggccggg tacggccgtt ttacgttcaat ggttaccgc ggggtacgttta ttatggat 20880
cggttataaa tgattttgcgatccgatccgtt aatagggtat gttgcgttgc 20940
tacataccggc gaataaaatgg gatattataa tattcggtat gtagtgcgttgc gtcgttgc 21000
atgttaacgaa agaaaatgtt tcgaaagaag gtttttttgc gttatgttgc gttttttataa 21060
aacaatggatgtt agcgatgttgcg tggtaaaaaat aacggaaatcat tcgttgcgttgc 21120
cggttataaaataatgtt ggttgcgttgc gtcgttgcgttgc ggggttgc gcaatgttgc 21180
atgcgtgttcc gtcggccgtt ttatggatgtt gttgcgttgc gtttttttgc gttttttataa 21240
aaatagatggg ttatccgtt catgcgttgc gtttttttgc gttatgttgc gttttttataa 21300

-continued

aattatcgtc gtattcgta tttgatatgt cgaatttcc gttaaaatta cgtggtaacgg 21360
 cggtaatgtc gttaaaagaa aatcaaataa atgatatgat atattcgta tttagaaaaag 21420
 gtcgttaat aatacgtgaa aataatcgta tagtagtata gtcggatata ttagtaaata 21480
 attaaacgaa catgtttata ttttattat ttttaacgtt aacgtcggt tcggatttag 21540
 atcgttgta gacgtttgat gatgtacaag cgccgaatta tacgcaacat acgtcgta 21600
 tgcgtgggt atattatccg gatgaaatat ttgcgttgcga tacgttatata ttaacgc 21660
 atttatttt accgtttat tcgaatgtaa cgggtttca tacgataaat catacgttt 21720
 gtaatccggt aataccgtt aaagatggta tatattttgc ggacggaa aaatcgaatg 21780
 tagtacgtgg ttgggtatcc ggttcgacga tgaataataa atcgcaatcg gtaataataa 21840
 taaataattc gacgaatgtta gtaatacgtg cgtgttaattt tgaattatgt gataatccgt 21900
 ttttgcgggt atcgaaaccg atgggtacgc aaacgcatac gatgatattt gataatcggt 21960
 ttaattgtac gtttgaatat atatcgatg cgttttcgtt agatgtatcg gaaaaatcgg 22020
 gtaattttaa acatttacgt gaatttgcgtt taaaaataaa agatgggttt ttatgtat 22080
 ataaagggtt tcaaccgata gatgttagtac gtgatttacc gtcgggttt aatacgtaa 22140
 aaccgatatt taaatttacgt ttaggtataa atataacgaa tttcgtgcg atattaacgg 22200
 cgtttgcgc ggcgcagat atatgggtt cgtcgccggc ggcgtatccc gtaggttattt 22260
 taaaaccgac gacgtttatg taaaaatatg atgaaaatgg tacgataacg gatgeggtag 22320
 attgttcgca aaatccgtt gccaattaa aatgttcggg aaaatcggtt gaaatagata 22380
 aagggtatata tcaaaccgtc aattttcgta tagtaccgtc ggggtatgtt gtacgtttc 22440
 cgaatataac gaatttgcgtt ccgtttgggt aagtatttta tgcgacggaa tttccgtcgg 22500
 tataatcggtt ggaacgtaaa aaaatatcgta attgtgtacg ggattattcg gtattatata 22560
 attcgacgtt ttttcgacg tttaaatgtt atggtgtatc ggcgacggaa ttaaatgatt 22620
 tatgttttc gaatgtatat gcccgttgcgtt tttttttttt aggtgtatgtt gtacgtcaaa 22680
 tagcgcggg tcaaaccgggt gtaatagcggtt attataatcg taaatttacgt gatgatattt 22740
 tgggttgtgtt attagcgtgg aatacgcgtt atatagatgc gacgtcgacg ggtattata 22800
 attataatcgta tcgttattta cgtcatggta aattacgtcc gtttgcgtt gatataatcg 22860
 atgttaccgtt ttcgcggat ggttaaccgt gtacgcggcc ggcgttaat ttatcggttgc 22920
 cgttaatgtt ttatgggtt tatacgacgtt cgggtatagg ttatcaacgg tttccgtcgg 22980
 tagtattatc gtttgcgtt ttaaatcggtt cggcgacgggt atgtggtccg aaattatcg 23040
 cggatttaat aaaaatcaa tttttttttt tttttttttt tttttttttt tttttttttt 23100
 tattaaacgcc gtcgtcgaaa cgtttcaac cgtttcaaca atttgcgtt gatgtatcg 23160
 attttacggta ttcggtaatcg gatccggaaa cgtcgaaat attagatata tccgggtgt 23220
 cgtttgggtgg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 23280
 tataatcgatgtt acggatgtat cgacggcgat acatgcggat caattaaacgc 23340
 cggcggtggcg tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 23400
 taggtcggtt acatgttagat acgtcgatgtt aatgtgtatc accgtatgtt gtcgggtat 23460
 gtgcgtcgta tcatacggtt tcgttattac gttcgacgtt gcaaaaatcg atagtagcgt 23520
 atacgtatgtt gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 23580
 cgacgtttttt ttcgtatgtt atacgtatgtt aatgtgtatc accgtatgtt gtcgggtat 23640

-continued

cggttagattt taatatgtat atatgtggtg attcgacgga atgtgcgaat ttattattac 23700
 aatatggttc gttttgtacg caattaaatc gtgcgttatac gggtatagcg gcggaaacaag 23760
 atcgtataac gcgtgaagta tttgcgcag taaaacaat gtataaaacg ccgaegttaa 23820
 aatatttgg tggtttaat tttcgcaaa tattaccgga tccgtaaaa ccgaecgaaac 23880
 gttcgtttat agaagatttta ttatthaata aagtaacgtt agcggatgct gggtttatga 23940
 aacaatatgg tgaatgttta ggtgatataa atgcgcgtga ttatataatgt ggcacaaaat 24000
 ttaatggttt aacggtatta ccggcgttat taacggatga tatgatagcg gcgtatacgg 24060
 cggcgttagt atcgggtacg gcgcacggcgg gttggacgtt tggtgcgggt gcggcgttac 24120
 aaataccgtt tgcgtatgcaa atggcgtatc gtttaatgg tataagggtta acgcaaaaatg 24180
 tattatatga aaatcaaaaa caaatagcga atcaatttaa taaagcgata tcgcaataac 24240
 aagaatcggtt aacgacgacg tcgacggcgt taggtaaatt acaagatgta gtaaatcaaa 24300
 atgcgcgaagc gttaaataacg ttagttaaac aattatcgtc gaattttggt gcgatatcgt 24360
 cggttataaa tgatataattt tcgcgttttag ataaagttaga agcggaaagta caaatagatc 24420
 gtttaataac gggtcgttta caatcggttac aaacgtatgt aacgcaacaa ttaatacgta 24480
 cggcggaaat acgtgcgtcg gcgaattttag cggcgcacgaa aatgtcggaa tgtgtattag 24540
 gtcaatcgaa acgtgttagat ttttgtggta aaggttatca ttatgtcg tttccgcag 24600
 cggcgcgcga tggtagtta ttttacatg taacgtatgt accgtcgcga gaacgttaatt 24660
 ttacgacggc gccggcgata tggcatgaa gtaaagcgta tttccgcgt gaagggttat 24720
 ttgttattaa tggtagtgcg tggttataa cgcaacgtaa ttttttcg ccgcaataaa 24780
 taacgacgga taatacgaaa gtatcggtt attgtgtatgt agtaataggt ataataaata 24840
 atacggata tgatccgtta caaccggaaat tagattcgtaa taaagaagaa tttagataat 24900
 attttaaaaa tcatacgtcg ccggatgttag atttaggtta tataatcggtt ataaatcggt 24960
 cggttagaaa tatacaaaaa gaaatagatc gtttaatgta agtagcggaa aatttaatg 25020
 aatcgtaat agatttacaa gaatttaggtt aatatgaaata atatataaaa tggccgttgt 25080
 atgtatggttt aggtttataa gcgggtttaa tagcgatagt aatggtaacg atattattat 25140
 gttgtatgac gtcgtgttgt tcgtgtttaa aagggtcgatg ttcgtgttgt tcgtgttgta 25200
 aattttagtgc agatgatgtcg gaaccggat taaaagggtgt aaaattacat tatacgtaaa 25260
 cgaacttatg gatttggta tgagatttt tactcttgcg tcaattactg cacagccagt 25320
 aaaaattgac aatgtttctc ctgcaagtac tggtcatgtc acagcaacgaa taccgctaca 25380
 agcctcaactc ctttcggat ggcttgcgtt tggcgatgtc tttcttgcgt ttttcagag 25440
 cgctaccaaa ataattgcgc tcaataaaaatg atggcageta gcccgttata agggcttcca 25500
 gttcatttgc aatttactgc tgctatttgtt taccatctat tcacatcttt tgcttgcgc 25560
 tgcaggatgtt gagggcgaat tttgtacatc ctatgccttg atatattttc tacaatgcgt 25620
 caacgcgtt agaatttata tgagatgttg gctttgttg aagtgcata ccaagaaccc 25680
 attactttt gatgcactt actttgttg ctggcacaca cataactatg actactgtat 25740
 accatataac agtgcacag atacaattgt cgttactgaa ggtgacggca tttcaacacc 25800
 aaaactcaaa gaagactacc aaattgggtgg ttattctgag gataggcact caggtgttaa 25860
 agactatgtc gttgtacatg gctatccac cgaagttac taccagtttgcgtt acgttacaca 25920
 aattactaca gacactggta ttgaaaatgc tacatttttc atcttaaca agcttgcataa 25980
 agacccacccg aatgtgcataa tacacacaat cgacggctct tcaggagtttgcgtt ctaatccac 26040

-continued

aatggatcca atttatgatg agccgacgac gactactagc gtgccttgc aagcacaaga 26100
 aagtgagtagc gaacttatgt actcattcgt ttcggaaagaa acaggtacgt taatagttaa 26160
 tagcgtactt cttttcttg cttdcggtt attcttgcta gtcacactag ccatacctac 26220
 tgcgcttcga ttgtgtgcgt actgtgc当地 tattgttaac gtgagtttag taaaaccaac 26280
 ggtttacgtc tactcgcgtg taaaaatct gaactcttct gaaggagttc ctgatcttct 26340
 ggtctaaacg aactaactat tattatttt ctgtttggaa cttaacatt gcttatcatg 26400
 gggataatg gtacgataac ggtagaagaa taaaacaat tattagaaca atggaattta 26460
 gtaatagttt ttttattttt agcgtggata atgttattac aatttgcgtt ttcgaatcgt 26520
 aatcgaaaa tatataataat aaaatttagta tttttatggt tattatggcc ggttaacgtt 26580
 ggtgttttgc tattagcggc ggtatategt ataaattggg taacgggtgg tatacgata 26640
 gogatggcgt gtatagttagg ttaatgtgg ttatcgtatt ttgttagcgtc gtttgcgtt 26700
 ttgcgcgtt aacgcgttgcgtt gttttttttt aatccggaaa cgaatataatt attaaatgt 26760
 ccgttacgtg gtacgatagt aacgcgtccg ttaatggaaat cggaaattagt aataggtgc 26820
 gtaataatac gtggcattt acgtatggcg ggtcattcgt taggtcgttgc tgatataaaa 26880
 gatttaccga aagaataaac ggttagcgcacg tcgcgtacgt tattcgttataa taaaatttagt 26940
 gogtcgcaac gtgttaggtac ggattcgggt tttgcggcgtt ataatcgta tcgtataggt 27000
 aattataaat taaatacggta tcatgcgggt tcgaatgata atatagcgtt attagtacaa 27060
 taagtgcacaa cagatgttca atcttgcgttgc cttccaggtt acaatagcag agatattgt 27120
 tatttttttttgc tatttttttttgc tatttttttttgc tatttttttttgc tatttttttttgc 27180
 agttagacaa ttatataacgcttgc ttttttttttgc ttttttttttgc ttttttttttgc 27240
 acctatggag ttagatttac cataaaacgta acatgaaaat tatttttttttgc ttttttttttgc 27300
 ttgttatttac atcttgcgttgc ttttttttttgc ttttttttttgc ttttttttttgc 27360
 tactaaaaga accttgcgttgc ttttttttttgc ttttttttttgc ttttttttttgc 27420
 ctgacaataaaat ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 27480
 gtactcgaca tacctatcgttgc ttttttttttgc ttttttttttgc ttttttttttgc 27540
 aagaggagggt tcaacaagag ctttttttttgc ttttttttttgc ttttttttttgc 27600
 tttaataact ttgttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 27660
 cttctatgggtt ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 27720
 ttgggttttca ctcgaaatcc aggatctaga agaaccttgc ttttttttttgc ttttttttttgc 27780
 gaaaacttctc attgttttgc ttttttttttgc ttttttttttgc ttttttttttgc 27840
 ggcgttgcgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 27900
 gtaataacttta tagcacttgc ttttttttttgc ttttttttttgc ttttttttttgc 27960
 ggcacactat gtttcaaaca ttttttttttgc ttttttttttgc ttttttttttgc 28020
 gtgggttgcgttgc ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 28080
 gagacgtact ttttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 28140
 tcaaaaccaac gtagtgcgttgc ttttttttttgc ttttttttttgc ttttttttttgc 28200
 aaccagaatg gaggacgcaag ttttttttttgc ttttttttttgc ttttttttttgc 28260
 aataataactg cgttttttttgc ttttttttttgc ttttttttttgc ttttttttttgc 28320
 cttcgaggccc aggccgttcc aatcaacacc aatagtggttgc cagatgacca aattggctac 28380

-continued

taccgaagag ctacccgacg agttcgtggt ggtgacggca aaatgaaaga gctcagcccc 28440
agatggtact tctattactt aggaactggc ccagaagett cacttcccta cggcgctaac 28500
aaagaaggca tcgtatgggt tgcaacttag gggagcccttga atacacccaa agaccacatt 28560
ggcacccgca atcctaataa caatgctgcc accegtgtac aacttcctca aggaacaaca 28620
ttgccaaaag gtttctacgc agagggaaagc agaggccggca gtcaagectc ttctcgctcc 28680
tcatcacgtt gtcgcgttaa ttcaagaaat tcaacttcctg gcagcagtag gggaaattct 28740
cctgctegaa tggctagccg aggtgggtaa actgcccctcg cgctattgct gctagacaga 28800
ttgaaccaggc ttgagagcaa agtttctggt aaaggccaac aacaacaagg ccaaactgtc 28860
actaagaaat ctgctgctga ggcacatctaaa aagcctcgcc aaaaacgtac tgccacaaaa 28920
cagtacaacg tcactcaagc atttgggaga cgtgggtccag aacaacccca aggaaatttc 28980
ggggaccacaa acctaatacag acaaggaact gattacaaac attggccgca aattgcacaa 29040
tttgctccaa gtgcctctgc attctttgga atgtcacgca ttggcatgga agtcacacct 29100
tcgggaacat ggctgactta tcatggagcc attaaattgg atgacaaaga tccacaattc 29160
aaagacaacg tcatactgtc gaacaagcac attgacgcat aaaaaacattt cccaccaaca 29220
gagcctaaaa aggacaaaaaa gaaaaagact gatgaagetc agcctttgcc gcagagacaa 29280
aagaagcagc ccactgtgac tcttcttcctt gcccgtgaca tggatgattt ctccagacaa 29340
cttcaaaattt ccatgagtgg agcttctgtc gattcaactc aggcataaac actcatgtat 29400
accacacaag gcagatggc tatgtaaacg ttttcgcaat tccgtttacg atacatagtc 29460
tactcttgcg cagaatgaat tctcgtaact aaacagcaca agtaggttta gttaaacttta 29520
atctcacata gcaatcttta atcaatgtt aacattaggg aggactgaa agagccacca 29580
cattttcatac gaggccacgc ggagtacgat cgagggtaca gtgaataatg ctaggagag 29640
ctgcctatat ggaagagccc taatgtgtaa attaatttt agtagtgcata tccccatgtg 29700
attttaatag cttcttagga gaatgacaaa aaaaaaaaaa aaaaaaaaaa a 29751

```
<210> SEQ ID NO 42
<211> LENGTH: 29751
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: synthetic
```

<400> SEQUENCE: 42

atatttaggtt ttacccatc caggaaaagc caaccaacct cgatctttg tagatctgtt	60
ctctaaacgc actttaaaat ctgtgttagct gtcgctcgcc tgcatgccta gtgcacccatc	120
gcagtataaa caataataaa ttttactgtc gttgacaaga aacgagtaac tcgtccctct	180
tctgcagact gcttacgggt tcgtccgtgt tgcatcgat catcagcata ccttaggttc	240
gtccgggtgt gaccgaaagg taagatggag tcattagtgtaggcgtttaa cgaaaaaaact	300
cacgttcaat tgtcactacc agtgttacag gtttagagacg tactcgtag ggggttcgggt	360
gattcagtgc aagaggact atccgaagct agagagacat ttaaaaacgg tacatgcgg	420
ttagtcgaac tcgaaaaagg cgtactacca caattggagc aaccatacgt ttttattaaa	480
cggtctgacg cacttagtac taatcacggt cataaggtag tcgagttagt cgccgaaatg	540
gacggatttc aatacggtag gtcaggtatt acactcgag tgtagtgcacacgttaggc	600
gaaacaccta tcgcttatcg taacgttcta ttgcgtaaaa acggtataaa gggcgcaggc	660
qqacattcat acqqtatcqa tcttaaqtcg tacgatttag qcqacqaact cqgtacqgtat	720

-continued

ccaatcgaaag attacgaca gaattggaa actaaacacg gatctggcgc attacgcgaa 780
cttacacgcg aacttaacgg aggccgcagtg actagatacg tcgataataa tttttcggt 840
ccagacggat atccactcgta ttgttataag gattttcgcg ctaggccagg taagtctatg 900
tgtacactta gogaacaact cgattatatac gaatctaaaa gagggcgtata ttgttgcgc 960
gatcacgaac acgaaaatcgc ttggttact gagcgatctg ataagtcata cgaacatcag 1020
actccattcg aaattaagtgc tgctaaaaa ttgcataactt ttaaaggcga atgtcctaaa 1080
ttcgccccactc cacttaactc taaggtaag gtattcaac ctagagtcga aaaaaaaaaa 1140
actgagggtt ttatggtag gattaggtca gtgtatccag tcgctagtcc acaagagtgt 1200
aataatatgc atctatctac acttatgaaa tgtaatcatt gtgacgaagt gtcatggcaa 1260
acatgcgatt tccttaaggc tacatgcgaa cattgcggta ccgaaaatct cgtaatcgag 1320
ggacctacta catgcggata cttacctact aacgcgcgtcg ttaaaatgcc atgtccagct 1380
tgtcaggatc cagagatagg tccagagcat agcggtgcg attatcataa tcattctaat 1440
atcgaaacta gattgcgtaa aggccgttagg acacgttgc ttccggatg cgtattcgca 1500
tacgttaggtt gttataataa gagagcgat tgggtgccta gagctagtgc cgatatcggt 1560
agcggacata ccggattac aggcgataac gttgagacac ttaacgagga tctgttagag 1620
attctatcac gogaacgcgt taatattaac atagtccgc attttcatct taacgaagag 1680
gtcgctatta tactcgctag ttttccgcg agtacatccg cttttatcg tactattaag 1740
tctcttgcact ataagtcttt taaaactatc gttgagtcat gcggtaatataa taaggttact 1800
aagggttaagc cagtggaaagg cgcatggaaat atcggtcagc aacggtcagt gcttacacca 1860
ctatgcgggtt ttccctgtca agccgcaggc gtaattagat ctatttcgc acgtacaccc 1920
gacgctgcata atcattctat tcccgcatacgt caacgtgcgt ccgttacgat actcgacgg 1980
attagcgaac agtcacttag actcggtgc gctatgggtt atacatccga tctgttaacg 2040
aatagtgtga ttattatggc ttacgttaca ggccgattag tgcaacacagac tagtcaatgg 2100
ttgtctaatc tgtaggtac tacagtcgaa aaattgcgc acatccgcgat atggattgg 2160
gctaaattgt ctgccccggc cgaattccctt aaagacgctt gggagataact taaatttcgt 2220
attaccggcg tattcgatata cgttaagggtt cagattcagg tcgctagcga taatattaag 2280
gattgcgtta agtgtttat cgacgttagt aataaggcgc tcgaaatgtg tatcgatcag 2340
gttacaatttgcg cccggagctaa gcttagatca cttaaacttag gcaaggtttt tatcgctcaa 2400
tctaagggtt tgatcgatca atgtatacgc ggtaaggagc aattgcattt gcttacgc 2460
cttaaggcgcctt ctaaaagaggt tacattccctt gagggcgcatt cacacgatac agtgcattact 2520
agcgaagagg ttgtgtttaa aaacggcgaa ctcgaagcac ttgagacacc agtgcattct 2580
tttactaactc ggcacatcgtt aggttacaccc gtagtgcgtt acggctttat gttgttgc 2640
attaaggata aagagcaata ttgcgcactt agtccaggc tgtagcgcac taataacgtt 2700
tttagactta agggaggcgc accttataaa ggccgttacat tcggtgagga tacagtttgg 2760
gaagtgcac ggtataaaaaa cgttaggattt acattcgac tcgacgcac tgctcgataag 2820
gtacttaacg aaaagtgttag cgttatataca gtcgaatccg gtactgaggt tactgagttc 2880
gcacgcgttag tcgctgaggc acgtgcgtt acattgcac ccgttacgc tctgttaact 2940
aatatgggtt tcgatcttgc cgaatggcgtca gtcgcactt tttatctattt cgtacgcgtt 3000
ggcgaagaga attttcgtc acgtatgtat tgttctttt accctcttgc cgaagaggaa 3060

-continued

gaggacgacg	ctgaatgcga	agaggaagag	atagacgaaa	catgcgaaca	cgaatacggt	3120
acggaagacg	attatcaggg	attgccactc	gaattcgccg	ctagcgctga	gactgttaga	3180
gtcgaagagg	aagaggaaga	ggattggta	gacgatacta	ctgagcaatc	cgaaatcgaa	3240
cccgaaaccg	aacctacacc	tgaggaaccc	gttaatcaat	ttacagggtt	tcttaagctt	3300
accgataacg	ttgcgattaa	atcgctttag	atcgatcaaag	aggctcaatc	cgctaattcct	3360
atggtaatcg	ttaacgctgc	taataatacat	cttaagcatg	gcccggccgt	tgcaggcgca	3420
cttaataagg	ctactaacgg	cgctatgcaa	aaagagtctg	acgattatata	taagcttaac	3480
ggtccactta	ccgtaggccg	atcatgtcta	ttgtcaggct	ataatctcg	taaaaaatgt	3540
cttcacgtag	tcggacctaa	tcttaacgct	ggcgaagaca	tacaattgct	taaagccgca	3600
tacgagaatt	ttaattcgca	agacatattt	ctcgacccac	tgttatccgc	aggtatttc	3660
ggagctaaac	cattgcaatc	attgcaggta	tgcgttcaga	cagtgcgtac	acaggatata	3720
atcgcagttt	acgataaggc	actatacgaa	caggtcgat	tggattatct	cgataatctt	3780
aaggcctagag	tcgaagctcc	taaacaggaa	gagccaccta	atactgagga	ttcttaagact	3840
gaggaaaaat	ccgttgtgca	aaaaccagtc	gacgttaagc	ctaaaattaa	ggcatgtatt	3900
gacgaagtga	ctactacact	tgaggagact	aagtttcttta	ctaataagtt	actgttattc	3960
gtgtatatta	acggtaagtt	gtatcacgt	tcacagaata	tgcttagagg	cgaagatatg	4020
tcttttctcg	aaaaagacgc	tccatatatg	gtaggcgacg	ttattacatc	tggcgatatt	4080
acatgcgtag	tgatacctag	taaaaaagcc	ggaggtacta	ccgaaatgct	atcacgtgca	4140
ctaaaaaaag	tgccagtcga	cgaatacatt	actacttac	ccggtcaggg	atgcgcaggt	4200
tatacactcg	aagaggctaa	aaccgcactt	aaaaaatgta	aatccgctt	ttacgttctg	4260
ccatctgagg	cacctaacgc	taaagaggag	atactcggtt	cagtgtcatg	gaattacgc	4320
gaaatgcttg	cgcacgtccg	agagcacacgt	aagcttatgc	ctattttat	ggacgtagg	4380
gctattatgg	ctactattca	acgtaagtt	aagggttata	agattcaaga	gggtatagtc	4440
gattacggag	tgagatttt	tttttataca	tctaaagagc	cagtcgttag	tattattact	4500
aagcttaact	cacttaacga	accatttagt	actatgccta	tccggatacgt	tacacacgg	4560
ttaatctcg	aagaggccgc	tagatgtat	cgatcactt	aggcaccacg	cgtatgtca	4620
gtgtcatcac	ctgacgcccgt	tactacttat	aacggatacc	ttacatctag	ttcttaagact	4680
agtgaggaac	attcggttga	gacagtgtca	cttgcgggtt	catatcgca	ttggtcttt	4740
tcaggtaac	gtactgagtt	aggcggttag	tttcttaaac	gcccggataa	gatagtgtat	4800
catacacttg	agtccccgt	cgagtttac	ttagacggcg	aagtgtttag	tctcgataag	4860
cttaagtctc	tattgtca	tagaggggtt	aagacttata	aagtgttac	tacagtgcgt	4920
aatactaata	tgcatacaca	gttagtcgt	atgtctatga	catacggtca	gcaattcgga	4980
cctacttata	ttgacgggtc	agacgttact	aaagattaagc	ctcacgttaa	tcacgaaggt	5040
aagactttt	tcgtatttgc	atctgacgt	acacttagat	ccgaaggatt	cgaataactat	5100
catacacttg	acgaatctt	tctcggttag	tatatgtca	cacttaacca	tacaaaaaaa	5160
tggaaattcc	cacaggttagg	cgggacttaca	tctattaaat	ggggcgataa	taattgttat	5220
ctgtcatcg	tgttactcgc	attgcagca	ctcgaagtga	aatttaacgc	tcctgcattt	5280
caagaggcat	actatagggc	tagagccgt	gacgctgat	atttttgcgc	acttataactc	5340
gcataactcta	ataaaaccgt	aggcgaactc	ggtgacgtta	gagagactat	gacacatcta	5400
ttgcaacacg	ctaatctcg	atccgctaaa	agagtgttta	acgtatgtg	taaacatgc	5460

US 9,476,032 B2

259

260

-continued

ggtcagaaaa ctactacact taccggagtc gaagccgtta tgtatatggg tacactatca 5520
tacgataatc ttaaaaccgg agtgtcaatt ccatgcgtat gggtaggga cgctacacaa 5580
taaccttagtc aacaagagtc tagttcggt atgatgtcg ctccacctgc cgaatacaaa 5640
ttgcaacagg gtacattctt atgcgctaac gagtataccg gtaattatca atgcggtcat 5700
tatacacata ttaccgctaa agagacattg tatacgatag acgggcaca tcctactaaa 5760
atgtctgagt ataagggtcc agttaactgac gtttttaca aagagactag ttatactact 5820
actattaaac cggttagtta taagttagac ggagtgactt ataccgaaat cgaacctaag 5880
ttagacggat actataaaaa agataacgct tattataccg aacagccaat cgatttagtg 5940
cctacacaaac cattgectaa cgctagttc gataattta agttacatg ttctaatact 6000
aaattcgcag acgatctta tcagatgacc ggtttacta agoctgctag tagagagtt 6060
tccgttactt ttttccgta tcttaacggt gacgttagtgc caatcgatta tcgctattac 6120
tctgctagtt taaaaaagg cgctaagcta ttgcataaac ctatcgttt gcatattaaat 6180
caggctacta ctaagactac tttaaacct aatacttggt gtcttagatg tccttggct 6240
actaaggccag tcgatacatc taattcattc gaagtgcgtt cagtcgagga tacacagggt 6300
atggataatc tcgcatgcga atcgcaacaa cctacatccg aagaggttgt cgagaatcc 6360
acaattcaga aagaggtaat cgaatgcgac gttaaagacta ccgaagtggg aggttaacggtt 6420
atacttaaac cgtctgacga aggccgttaag gttacgcaag agttaggcga tgaggatctt 6480
atggccgcattt acgttgagaa tacatctatt actattaaaa aacctaacga actatcactc 6540
gcatttaggtt ttaagactat cgctacacac ggtatacgcc ctattaaatc gggttccatgg 6600
tctaagataac tcgcatacgt taagccttt ctggcgtcaag ccgttattac tacatctaat 6660
tgcgctaaac gacttgcgca acgcgtttt aataattata tgccatacgt ttacatttg 6720
cttttcaat tttttttactt tactaagtct actaactcag gtatacgcc tagtctacct 6780
actactatcg ctaaaaatcc cgttaaatcc gttgcgaaat tttttttactt tttttttactt 6840
aattacgtt agtcacctaa attttctaaa ttgtttacaa tcgctatgtg tttttttactt 6900
ttgtctattt gtctcggttc attgatttc gttaccgcgtt cattcggtt gttttttactt 6960
aatttcggcg cacctagtta ttgttaacggg gttagagagt tttttttactt tttttttactt 7020
gttactacta tggatttttt cgaaggatct tttttttactt tttttttactt tttttttactt 7080
gattcactcg attcatatcc cgtctcgaa actatttcagg ttacgatttag ttcttataaa 7140
ctcgatctta ctatactcgg tctagccgtt gaatgggtgt tagcgtatataatc 7200
aagttttttt acttactcgg tctatccgtt attatgcgtt tttttttactt tttttttactt 7260
agtcattttt tttttttactt tttttttactt tttttttactt tttttttactt tttttttactt 7320
ccagttacgtt ctaggttacgtt tttttttactt tttttttactt tttttttactt tttttttactt 7380
tcatacggtt atttatggg cggatgtaca tctagttatc gtatgtgtt ttataaacgtt 7440
aatcgcgttta ctagagtcgtt tttttttactt tttttttactt tttttttactt tttttttactt 7500
gtttacgttta acggaggttag ggggtttgtt aagactcataatc 7560
gatactttttt tttttttactt tttttttactt tttttttactt tttttttactt tttttttactt 7620
caatttaaac gtccaaattaa ccctaccgtt caatctgtt atatagtcgtt tagcgttgcgtt 7680
gtttttttactt tttttttactt tttttttactt tttttttactt tttttttactt tttttttactt 7740
catccactat cacatttcgtt taacttagac aatcttaggg ctaataatac taagggttgtt 7800

-continued

ctgccaatta acgtaatcgt attcgacggt aagtctaaat gcgacgaatc cgctagtaag	7860
tctgctatgt tgtattactc tcagttatg tgtcaaccta tactgttact cgatcaggca	7920
ttagtgtcag acgttaggcga tagtactgag gttagcgtta aatgttcga cgcttacgtc	7980
gatacttttgcgctacatt ctcagtgcct atggagaaac ttaaggcatt agtcgctacc	8040
gctcattctg agttagcgaa aggcggtcg tttagacggag tgctatctac attcgtatct	8100
gcccgttagac aggccgttagt cgatactgac gtgcatacta aagacgtaat cgaatgtctt	8160
aagctatcac atcattccga tcttgagggtt acaggcgatt catgtataaa ttttatgttt	8220
acatataata aggtcgagaa tatgacaccc agagacttag ggcgcgttat cgattgtaac	8280
gctagacata ttaacgctca gggtgcgaaa tcacataacg tatcattgtat ttggAACGTT	8340
aaagactata tgtcaactatc tgagcaattt cgtaaacaga tacgatccgc tgctaaaaaa	8400
aataatatac cgtttagact tacatgcgt actactagac aggtcgtaa cgttattact	8460
actaagatta gtcttaaggg aggttaagatc gtttagtacat gtttaagct tatgttaag	8520
gtacactgt tatgcgtact cgctgcactc gtatgttata tcgtttagcc agtgcataca	8580
cstatctattc acgacggata tactaacgaa attatcgat ataaggctat tcaagacgga	8640
gtgacacgcg atattattatg tactgacgt tgttcgcta ataaacacgc tggattcgac	8700
gcttggttt ccgcaacgcgg agggtcataat aaaaacgata agtcatgtcc agtcgttgcc	8760
gctattattatc cacgcgaaat cggattcata gtgccagggt taccgggtac agtgggtcggt	8820
gctattaacg gcgatttttgcattttctg ccacgcgtt tttagcgcagt cggtatatt	8880
tgttatacac ctagtaagtt aatcgaataac tctgatttcg ctactagcgc atgcgtactt	8940
gccgctgagt gtactatttt taaagacgct atgggttaagc cagtgcata ttgttacgat	9000
actaatctgt tagagggtac tatttcatac tctgagttgc gacctgatac taggtacgta	9060
cttatggacg gatctattat tcaattccct aatacttac ttgagggatc cgtagagtc	9120
gttactacat tcgacgctga gtattgtaga cacggatcat gcaacgatc tgaggcggt	9180
atttgtctat ctacatctgg tagatggta cttataacg aacattatcg cgcaactatct	9240
ggcggtttttt gcccgttaga cgctatgaa ctatcgcta atattttac accatttagtgc	9300
caaccagtcg ggcacccgtga cgtatccgtc agcgtatgcg caggcggtat tattcgatata	9360
ctcggttacat ggcgtcgata ctatccgtt aaatttcgtt gggtattccgg tgagtataat	9420
cacgtatcg ctgctaacgc tctatttttcc cttatgttctt ttacgatact gtgtctcgta	9480
cctgcataact cttttctgac aggcgtatac tctgtgtttt atctgtatct tacatccat	9540
tttactaacg acgtttcgat tctcgacat ctgcaatggt tgcgtatgtt ttcacccata	9600
gtgccatttt ggattaccgc tatttacgtt tttgttattt cgcttaagca ttgtcatgtt	9660
tttttaata attatccgt taaacgcgtt atgtttaaacg gagtgacttt ttcgacattc	9720
gaagaggccg cactttgtac ttttctgtt aataaaagaga tggatcttaa gttacgggtct	9780
gagacactgt taccacttac acaataacaat aggtatctcg cactatataaa taagtataag	9840
tatccccatcg ggcacccgtga tactacatct tattcgatgc cgcgtatgtt tcatctcgat	9900
aaagcgctta acgatccgtt taaactctggc gcagacgtac tggatcaacc tccacaaaca	9960
tctattacgt ctgcccgtact gcaatccgtt tttcgtaaaa tggctttcc atcaggttaag	10020
gtcgagggtatgttgatgc agttacatgc ggtactacta cacttaacgg tctatggtaa	10080
gacgatacgg tttttgtcc tagacacgtt atttgtactg ccgaagatat gcttaaccct	10140
aattacgaag acttactgtat acgttaatct aatcactctt ttttagtgca agccggtaac	10200

-continued

gttcaattgc gcgttatcg tcattctatg caaaattgtc tgtaacgtct taaagtcat 10260
 acatctaatac ctaaaaactcc taagtataaa ttcgttagga ttcaacccgg tcagacattc 10320
 tcagtaactcg catgttataa cggttcaccc agcggagtgt atcaatgcgc tatgcaccc 10380
 aatcatacaa ttaaggatc ttttcttaac ggttcatgcg gatcagtccg ttttaatatc 10440
 gattacgatt gcgttagttt ttgttatatg catcatatgg agttacctac aggcgtacac 10500
 gctggactg atcttgaggg taagttttac ggtccattcg ttgatcgta aaccgctcaa 10560
 gccgcaggta ccgataactac tattacaatt aacgtactcg catggttata cgctgccgtt 10620
 attaacggcg atagatggtt tcttaataga tttaactacta cacttaacga ttttaactta 10680
 gtcgctatga aatataatta cgaaccactt acacaggatc acgtcgatat actcggtcca 10740
 cttagtgcgc aaaccggtat agccgtactc gatatgtcgct ctgcacttaa agagttactg 10800
 caaaacggta tgaacggtag gactatactc ggttagtacta tactcgaaga cgaatttaca 10860
 ccatcgcacg tcgttagaca gtgttcaggg gttacatttc agggtaagtt taaaaaaatc 10920
 gttaagggtt cacatcattt gatgttgctt acattcctta ctagtctatt gatactcgta 10980
 caatctacac aatggtcatt gtttttttc gtatacggaa acgcccccc gccttttaca 11040
 ctccggattt tggctatagc cgcattgcgt atgttggtag tggaaacataa acacgcattc 11100
 ttatgtctat tcctattgcc tagtctcgat acagtcgcatt actttaatat ggtgtatgt 11160
 cctgcttagtt gggttatgcg tattatgact tggcttgagt tagccgatc atctctatct 11220
 ggttatagac ttaaggattt cggttatgtac gcttagtcat tagtgttact gatactttag 11280
 actgcacgta cagtttacga cgacgctgtc agacgcgtt ggacacttat gaacgttatt 11340
 acacttgtgt ataaagtgtt ttacggtaac gctctcgatc aggcttattag tatgtgggca 11400
 tttagtgcata cgcgttacgtc taattactct ggagtcgttta cgactattat gtttctcgct 11460
 agggctatcg tattgtatg cggttagtact tatccactat tgtttattac cggtatata 11520
 ttgcaatgtt ttagtgcata gttttttttt tttaggtattt gttgttgcgtt ttatccgga 11580
 ttgtttgtc tacttaatag gtattttaga cttacattag gcttatacga ttatctcggt 11640
 agtacgcaag agtttagata tatgaattca cagggactgt taccaccaa atcttcaatc 11700
 gacgcttttta agcttaacat taatgttactc ggtataggcg gtaagecttg tattaagggtt 11760
 gcgacagtgc aatctaaat gtctgacgtt aagtgtacta gtgtcgact gttatcgtt 11820
 ttacagcaat tgcgagtcga atctgttctt aaattgtggg ctcaatgcgt tcaattgcatt 11880
 aacgatatact tactcgctaa ggataactact gaggcattcg aaaaaatgggt tagtctgtt 11940
 agcgtactgt tatctatgca aggcgcagtc gatattaatc gattatgcga agagatgtt 12000
 gacaatagggtt ctacattgcg gggttattcg tcagagttt ctatgttgcctt atcatacgct 12060
 gcatacgcata cccgtcaaga ggcatacgaa caggcgtcg ctacggcgaa ttctgaggta 12120
 gtgcttaaaa aacttaaaaa atcacttaac gttgcgaaat ctgagttcgat tagagacgct 12180
 gctatgcac gtaaaactcgaa aaaaatggcc gatcaggctt tgacacaaat gtataaacag 12240
 gctatgtctg aggataagcg tgctaaagggtt acatccgatc tgctaaactat gttgttact 12300
 atgttgcgtt aactcgatcc cgcacgcattt aacaatattt ttaataacgc tagagacggtt 12360
 tgcgttaccac ttaatattat accgttactt actgcgttgcgta agctttaggtt agtgcgttccc 12420
 gattacggta cttataagaa tacatgcgcac ggtaataactt ttacatacgc tagtgcattt 12480
 tgggagattc aacaggtcgat tgacgctgtat tctaaaatcg ttcaattgtc tgagatata 12540

-continued

atggataact cacctaatactc cgcatggcca ttaatcgta cagcgtaacg cgctaactct 12600
gccgttaagt tacagaataaa cgaattgtca ccagtcgcatt tgcgtaaat gtcatgcgc 12660
gcccgtacta cacaaccgc ttgtactgac gataacgctc tcgcataacta taataactct 12720
aaaggcggta gatcgtaact cgcaacttata tccgatcatc aggatctaa atgggctaga 12780
ttccctaagt ctgacggtaac aggtacaatt tataccgaac tcgaaccacc atgttagattc 12840
gttaccgata cacctaaggaa acctaaggaa aagtatctat actttattaa gggactgaat 12900
aatctgaata ggggtatggt gttagggtca ctgcagcta ccgttagatt gcaagccgt 12960
aacgctactg aggttccgc taactcaacc gtacttagtt tttgcgcatt cgcaactcgat 13020
ccagctaagg cttataagga ttacttagct agcggaggtaa aacctattac taattgcgtt 13080
aaaatgtat gtacacatac cggtacaggt caggctattaa cggttacacc tgaggctaat 13140
atggatcagg agtcattcg aggccatca tggcgatctat attgttagatg tcataatcgat 13200
catcctaatac ctaagggttt ttgcgatctt aagggttaatg acgttcagat tcctactaca 13260
tgcgctaacg atccagtcgg ttttacattt cgtaataccg tttgtacatg atgcggatgt 13320
tgaaaaggaa acggatgttc atgcgatcag ttacgcgaac cacttatgca atccgctgac 13380
gtatgtatcat tccttaatgg gtttgcgggtg taagtgcgcg tagattgact ccatgcggta 13440
caggtacttag tactgacgta gtgtatcgcg cattcgatatt ttataacgaa aagggttgcgc 13500
gattcgctaa attccttaag actaattgtt gtagattca ggaaaaagac gaagagggtta 13560
atctgttaga ctcatatcc gtagtggaaac gtcatactat gtctaattat caacacgaa 13620
agactattta taatctcgat aaggattgtc cagccgttgc ggtacacgat ttttttaat 13680
tttagagtcga cggtgatatg gtgccacata tttctagaca gagattgact aagtataacta 13740
tggccgatct agtttacgct cttagacatt tcgacgaaagg taatttgcgtt acacttaaag 13800
agataactcgat tacatataat tggcgacg acgattactt taataaaaaa gattggatcg 13860
atttcggtta gaatccccat atacttaggg ttacgctaa cttaggcgaa cgcgttagac 13920
agtcattcgat taagacagtg caattttgcg acgttatcgat tgacgcaggat atagtcggcg 13980
tacttacact cgataatcag gatcttaacg gtaatttgcgtt cgatttgcgcg gatttcgttc 14040
aggtcgacc aggttgcggaa gtgcctatag tcgattcata ctattcatg cttatgccta 14100
tacttacact tacacgtgca cttgcgtgtg agtcacatattt ggacgcgtat ctcgtaaac 14160
cattgattaa atgggatcta cttaaatacg attttactga ggaacgatttgcgttattcg 14220
ataggatattt taagtattgg gtcacgactt atcatcttaa ttgttataat tgcgttgcgtt 14280
atagatgtat actgcattgc gctaattttta acgtattgtt ctcaaccgta ttcccaccta 14340
catcattcggtt cccactcgta cgtaagattt tcggttgcgtt cgtaccatc gtagtgcata 14400
ccggatatca ttttagagag tttaggcgtt tgccataatca ggacgttaat ctgcattcta 14460
gttagactatc tttaaagag ttgttgcgtt acgttgcggaa tccagctatg catgcgttca 14520
ggcgtaatct gttactcgat aagggtacta catgttttag cggttgcgcata cttactaata 14580
acgttgcgtt tcagactgtt aaaccggta attttataaa ggattttac gatttcgttgc 14640
tgagtaaggg tttttttaaa gagggatcta gcgttgcgtt taagcattttt tttttcgctc 14700
aagacggtaa cgctgttattt agcgattacg attactatag atataatctg ccaactatgt 14760
gcatatcgat tcaatttgcgtt ttgcgttgcgtt aggtatcgatg taagtatttc gatttgcgtt 14820
acggagggtt gatataacgct aatcaggtaa tcgttataaa tctcgataag tctgcggat 14880
tcccatcaa taaatgggtt aaggcttagat tgttattacgta ttctatgtca tacgaggatc 14940

-continued

aggacgtctt attcgcataat actaagcgta acgttataacc gacaattacg caaatgaatc 15000
ttaaatacgc tattagcgct aaaaatcggt caegtacagt cgccggcgta tcaatttgta 15060
gtactatgac taatcgtaa ttccacaaaa aattgcctaa gtctattgcc gctactagag 15120
gegctacagt cgttatcggt acatctaaat ttacggagg ttggcataat atgcttaaga 15180
cagtgatttc tgacgttag actccacatc ttatgggtt ggattaccct aatgcgata 15240
gggctatgcc taatatgtg cgtattatgg ctatgtcgat actcgacgt aaacataata 15300
catgttgtaa tctatcacat agattctata gactcgtaa cgaatcgct caagtgttta 15360
gegaaatggt tatgtgtggc ggatcactt acgttaagcc tggcggtaca tctageggag 15420
acgtactac cgcttacgct aattccgtt ttaatatttg tcaaggcgtt accgcgtaac 15480
ttaacgctctt attgtctact gacggtaata agattgccga taaatacgta cgtaatctgc 15540
aacatagatt gtacgaatgt ctgtatcgta atcgcatgt cgatcacgaa ttcgttgacg 15600
aattttacgc ttacttacgt aaacattttt ctatgtatgt tctatctgac gacgcagtcg 15660
tttggtataa ctctaattac gctgcacagg ggtagtcgc tagtattaag aattttaaag 15720
ccgtactgta ttatcagaat aacgtttta tgcgtggggc taagtgttgg accgaaacccg 15780
atctgactaa gggtccacac gaattttgtt cacagcatac tatgctcggtt aaacagggtg 15840
acgattacgt ttatctacca tatcccgtc cgtcacgtat attaggcga gggttggc 15900
ttgacgatat cgtaaagact gacggtacac ttatgttga gagattcggtt agtctcgcaa 15960
tcgacgctta tccattgact aagcatccta atcaggaaata cgctgacggtt ttctacttat 16020
acttacagta tatacgtaaa ttgcacgacg aacttacagg gcatatcggtt gatatgtatt 16080
cggttatgct tactaacgat aatactagta ggtattgggaa acctgagttt tacgaagcta 16140
tgtatacacc acataccgta ctgcaagccg taggcgcatg cgtactatgt aattcacaga 16200
ctagtcttag atgtggcgca tgcgtatcgta gaccattttt atgttgcgtt tgcgttacg 16260
atcacgttat tagtacatca cataagtttag tgcgtatcgtaaaccatc gtttgcgttacg 16320
ctccaggttg cgacgttact gacgttacgc aattgtactt aggccgtatg tcatactatt 16380
gttaaatcaca taaacctcctt attagtttc cgttatgcgc taacggtcag gtattcggtc 16440
tatataaagaa tacatcgtaa gggctcgata acgttaccga tttaacgctt atcgctacat 16500
gcgattggac taacgctggc gattatatac tgcgtatcgta atgtactgag cgacttaagt 16560
tattcgacg cggaaacactt aaggctaccg aagagactttt taaattgtca tacggatcg 16620
ctaccgtacg cgaagtgctt agcgatagag agttgcacatc atcatgggag gtcggtaagc 16680
ctagaccacc acttaatcgtaaattacgtttt ttacagggtt tgcgttactt aagaattcgta 16740
aagtgcataat tggcgaatat acattcgaaa aaggcgatca cggtgacgcc gtagtgcata 16800
gagggtactac tacttataaag cttaacgttag gcgattttt cgtacttaca tcacataccg 16860
ttatgccatt atccgctcctt acattagtc cacaagagca ttacgtttagg attaccggat 16920
tgtatcctac acttaatattt tctgcgtat ttcatctaa cgttgcgtt tatcagaataa 16980
tgggtatgca aaaatactctt acattcgagg gaccccccgg taccggtaag tctcatttcg 17040
caatcgactt cgcattgtat tattcgatgtt ctggattgtt gtagtacgct tgcgttacacg 17100
ctgcgttgc cgcgtatcgtaa gaaaagctt ttaagttatctt accaattcgat aagtgttcc 17160
gtattataacc cgcttagggctt agagtcgaat gtttcgatataa gttttaaagtg aattcgacac 17220
tcgaacaataa cgttttttgtt acgggttaacg ctctaccaggaa gactactgccc gatatcgtag 17280

-continued

tgttcgacga aattctatgt gctactaatt acgatctatc agtcgttaac gctagactta 17340
gggctaagca ttacgtttat ataggcgatc cagcgcaatt gcctgcacca cgtacattgt 17400
tgcataagg tacactcgaa ccttagtatt ttaattcagt gtgttagactt atgaaaacta 17460
teggacctga tatgtttctc ggtacatgtc gtagatgtcc tgccgaaata gtcgatacag 17520
ttagcgcact cgtatacgat aataagctt aggcacataa ggataagtct gcgcaatgtt 17580
ttaaaatgtt ttacaaggc gtaattacac acgacgttag ttccgcattt aatagaccac 17640
agataggcgt agtgagagag tttcttacac gtaatccccg ttggcgtaaa gccgtttta 17700
tttagtccata taactctcaa aacgcagtcg cttagtaagat actcggattt cctacacaga 17760
cagtcgattt tagtcagggt agcgaataacg attacgttat ttttacgcaa actaccgaaa 17820
ccgctcattt atgttaacggtt aatagattt acgttgcgt tactagggtt aaaatcggtt 17880
ttctatgttat tatgtctgtat cgcgatctat acgataagtt gcaatttaca tcacttgaga 17940
taccttagacg taacgttgcg acattgcaag ccgaaaacgt taccggattt tttaggattt 18000
gttctaaagat tattaccgga ttgcattcata cacaggcacc tacacatcta tcagtcgata 18060
ttaagtttaa gactgagggtt ctatgcgttg acatacccg tataacctaa gatattgactt 18120
atcgttagatt gatttagtattt atgggtttta aatgaattt tcaggtaac ggttatccta 18180
atatgtttat tacacgcgaa gaggcttata gacacgttag ggcttggata gggttcgacg 18240
ttgagggtatg tcacgctaca cgtgacgcag tcggtaactaa tctgccattt caattagggt 18300
tttcgacagg cgtaatctg gttgcgtac ctaccggata cgtcgataact gagaataata 18360
ctcgagtttac tagggtaac gctaaacctc cacctggcga tcaattttaa catctgattt 18420
cacttatgtttaaagggatg ccttggaaacg tagtgcgtat taagattgtt caaatgttta 18480
gcgatacact taagggatg tctgataggg ttgtgttctg actttggctt cacggattcg 18540
aattgacatc tatgaaatat ttgcgttaaa tcggacactga gegtacatgt tgcgttatcg 18600
ataagcgtgc tacatgtttt agtacatcta gcgatacata cgcttggat aatcattcag 18660
tcggattcga ttacgtttat aatcctttt tgattgacgt ttagcaatgg gggttaccg 18720
gtaatctgca atctaattcac gatcagcatt gtcaggtaa cggtaacgct cacgttgcgt 18780
catgcgcacgc tattatgact agatgtcttgc cctgacacgat atgttcgtt aagagatcg 18840
atgggtcgtt cgaataccctt attataggcg acgaacttag ggtaattcc gcttgcgtt 18900
aagtgcaca tattggcgat aagtctgcgtt tattagccga taagttccc gtattgcacg 18960
atatcggttaa tcctaaagcg attaaatgcg taccacaagc cgaagtcgaa tggaaatttt 19020
acgacgctca accatgttct gataaggcat acaaaatcgaa agagttgttt tactcatacg 19080
ctacacatca cgataagttt actgacggcg tatgtctgtt ttggaaattgtt aacgttgcata 19140
ggtagtcccgta acacgtatc gtttgttagat tcgataacttag gggttctatct aatctgaatc 19200
tgccagggttgc cgtatggcgat tcactatacg ttaataaaaca cgctttcat acacgtcgat 19260
tcgataaattc cgctttacg aatcttaagc aattgcctt ttttattac tctgatagtc 19320
catgcgaatc tcacggttaag caagtcgttgc gcgatatacgat ttacgttcca cttaaatccg 19380
ctacatgttat tacacgttgtt aatctcgag ggcgcgtttg tagacatcagc gctaacgaat 19440
atcgtcaata cttagacgct tataatatgat tgatttagtgc cgggtttca ttgtggattt 19500
ataagcaattt cgataactttt aatctatggat atacttttac tagattgaa tctctcgaaa 19560
acggtgcgtt taacgtcgat aataagggttgc atttcgacgg tcacgctggc gaagctcccg 19620
tataattat taataacgca gtgtataacta aggtgcacgg tattgcgttc gagattttcg 19680

-continued

aaaataagac tacattaccc gttaacgttg cattcgaatt gtgggctaaa cgtaatatta 19740
aaccagtgcc agagattaag atacttaata acttaggcgt tcatatcgct gctaataccg 19800
ttatggga ttataaacgc gaagctcctg cacatgtgtc aactataggc gtatgtacta 19860
tgactgatat cgctaaaaaa cctaccgaat ccgcttgc acacttaca gtgttattcg 19920
acggtaggggt tgagggtcag gtcgatctgt ttcgtAACGC acgttaacggc gtactgatta 19980
ctgagggatc ggttaaggga ttgacaccta gtaaggacc tgctcaagct agcgtaacg 20040
gagtgacact tataggcgaa tcagttaga ctcattaa ttactttaa aaagtgcacg 20100
gtattattca acagttacca gagacttatt ttacacaatc acgcgatctt gaggattta 20160
aacctagatc tcaaattggag actgatttc tcgaactcgc tatggacgaa ttattcaac 20220
gatataagct tgaggatac gcattcgaac atatcggtt cggcgatttt tcacacggtc 20280
agttaggccg attgcattttt atgatcggtc tagctaaacg atcacaggat agtccacta 20340
agcttggaga ttttataacct atggattcaa ccgttaagaa ttactttatt accgatgcgc 20400
aaaccgggtt atctaaatgc gttttagcg taatcgattt actgttagac gatttcgtt 20460
agattattaa gtctcaggat ctatcgtga ttagtaaggat agtgaaagtg acaatcgatt 20520
acgctgagat ttctttatg ctttgggtta aagacggtca cgtttagact tttacccta 20580
aattgcaggc tagtcaggca tggcaacctg gagtcgtat gcttaatttgc tataaaatgc 20640
aacgtatgtt actcgaaaaaa tgcgatctgc aaaattacgg tgagaacgca gtgataacta 20700
agggttattat gatgaacgtc gctaagtata cacaattgtt tcaataacctt aatacactt 20760
cacttgcgtt accatataat atgegagtga tacatttcgg agccggatct gataaggcg 20820
ttgcgcagg tactgcgtt ttgcgttacat ggtgccttac cggtagactt ttagtgcatt 20880
ccgatcttaa cgatttcgtt tctgacgtt atagtagactt tataaggcgat tgctacatcg 20940
tgcataaccgc taataaatgg gatctgatattt ttagtgcataat gtaatccca cgtactaaac 21000
acgttacgaa agagaacgtat tctaaagagg gtttttttac atatctatgc gtttttattt 21060
aacagaaact cgcattaggc ggtcttattt cctttaagat tactgagcat agttggaaacg 21120
ctgatctgtt taagttatg ggttattttt gttgggtggac cgcatttcgtt actaacgtt 21180
acgcttagttc tagcgaagca ttcttaatcg gggcttaattt tctcggttac cctaaagagc 21240
aaatcgacgg atatactatg cacgctaattt atattttttgc gcttaatactt aatccattt 21300
aattgtcttag ttatttcattt ttcgtatgtt ctaaatttcc acttaatgtt cgcggactt 21360
ccgttatgtc acttaaagag aatcagatattt acgatatgtt ttacttttta ctcgaaaagg 21420
gtaggttgcata tatacgcaaa aataataggg tagtcgttag ttctgacata ctcgttata 21480
ataaaacgaa catgtttttt tttctgttacat tccatccgtt tcagatctcg 21540
atagatgtac tacatttcgac gacgttcagg cacctaattt tacacagcat acatctgtt 21600
tgagaggcgtt atactatcctt gacgaaattt ttagatccgtt tacattgtt cttacacagg 21660
atctattctt accttttttac tcttaacgtt cagggtttca tacaattat catacattcg 21720
gtaatccgtt tataccgtt aaagacggta ttatccgtt acgttacgaa aatcttaacg 21780
tagttagagggtt ggtagtacta tgaataataa gtctcaatca gtgatttattt 21840
ttaataatttca gactaacgtt gtgtacgcg catgttattt cgaatttgc gataatcctt 21900
tttgcgtt gtcataaccc atgggtacac agactcatac ttttttttgc gataacgtt 21960
ttaattgttac attcgtatattt tttctgttac ctttttactt tgacgtatcc gaaaaatctg 22020

-continued

gtaatttaa acacttacgc gaattcggtt taaaaataa agacggttt ttgtacggtt 22080
 ataaggata ccaacccatc gacgtcgta gggatctgcg atcaggtttt aatacaccta 22140
 agcctattt taagttaccg ttaggcatta atattactaa ttttagagct atacttaccg 22200
 ctttagtcc agctcaggat atttgggtta cttagcgtgc cgcatatttc gtccggatatc 22260
 ttaaacccatc tactttatg cttaaatacg acgaaaacgg tacaattact gacgcagtcg 22320
 attgttcaca gaatccatta gccgaactta agtgttcagt taagtcatc gaaatcgata 22380
 agggtattta tcagactagt aatttttagag tcgtacctag cggtgacgta gtgagatcc 22440
 ctaatattac gaatctatgt ccattcgccg aagtgtttaa cgctactaaa ttccctagcg 22500
 tatacgcggg ggagcgtaaa aaaatttagta attgcgttgc cgattactct gtgttgata 22560
 attcgacttt ttttagtact tttaagtgtt acggagtgtc agctactaag cttaacgatc 22620
 tatgttttc taacgtatac gctgattcat tcgttagtgaa aggcgatgac gttaggcaaa 22680
 tcgctcccg tcagactggc gtaatcgctg attataatta taagttacct gacgatttta 22740
 tgggttgcgt actcgcattgg aatacacgta atatcgacgc tacttagtacc ggttaattata 22800
 attataagta tagatatctt agacacggta agcttagacc attcgaacgc gatatttcta 22860
 acgtaccatt ctcacctgac ggtaagcctt gtacacctcc tgcacttaat tgtttattggc 22920
 cacttaacga ttacggtttt tacactacta ccggtatagg gttcaacct tatagggtag 22980
 tcgttctatc attcgaattt cttaacgcgc cagctaccgt atgcggaccc aaattgtcaa 23040
 cccatctgtat taagaatcaa tgccgttaatt ttaattttaa ccgtcttacc ggtacaggcg 23100
 tacttacacc tagttctaaa ccgtttcaac ctttcagca attcggtagg gacgttagcg 23160
 attttaccga tagcgttagg gatcctaaaa ctatgtgagat actcgatatt agtccatgcf 23220
 cattccggagg cgtaagtgtg attactcccg gtactaacgc tagttctgag gttccgtac 23280
 tgtatcagga cgtaatttgt actgacgtat caaccgctat acacgctgtat caattgacac 23340
 ctgcgtggcg tatatactct accggtaata acgttttca gacacaagcc gggtgtctga 23400
 taggcgcaga gcatgtcgat acatcatacg aatgcgatataccgataggc gcaggtattt 23460
 ggcgttagtta tcatacagtg tcattgcgtt gatctactag tcagaaatca atcgtgcgt 23520
 atactatgtc attaggcgct gatagttcaa tcgcatactc taataatact atcgtatac 23580
 ctactaattt ttcaatttcg attactactg aggttatgcc agtgtctatg gctaaagacta 23640
 gtgtcgattt taatatgtat atttgtggcg attcaaccga atgcgtataat ctgttattgc 23700
 aatacggatc tttttgtacg caattgaatc gtgcactatc aggtatagcc gctgaacagg 23760
 atcgtataac tagagaggta ttgcacagg ttaaacagat gtataagact cctacactta 23820
 agtatttcgg agggtttaat tttcacaga ttttaccga tccacttaaa cctactaaac 23880
 gatctttat tgaggatctg ttattcaata aggttacact cgccagacgc ggttttatga 23940
 aacaatacgg cgaatgttta ggcgtatatta acgcttagggta tctgatttgc gctaaaaat 24000
 ttaacggctc tacagtgtta ccaccactat tgactgcgtatgatagcc gcatatactg 24060
 cccgattatgt gtcaggatcc gctaccgcag gttggacatt ccgtgcccgt gccgcattgc 24120
 agatccattt cgctatgcaa atggcttata gatthaacgg tataaggcggtt acgcaaaacg 24180
 tactttacga gaatcagaaa caaatcgcta accaatttaa taaggctatt agtcagattc 24240
 aagagtcaact tactactact agtaccgcac tcggtaagt gcaagacgtc gttatcaga 24300
 acgctcaggc acttaatacaca tcgcgttaaccg aattgtcttagt taatttcggtt gctatttagtt 24360
 cagtcgtttaa cgtatattctt tcttagactcg ataaggtcgaa agccgaaatgcg 24420

-continued

gattgattac cggttaggtg caatctcgc aaacatacgt tacgcaacaa ttgatacgcg	24480
ctgctgagat tagggctagc gctaattcg cagctactaa aatgtctgag tgcgtactcg	24540
gtcaatctaa aagagtcat tttgcggta aggggtatca tcttatgtct tttccacaag	24600
ccgcaccaca cggagtggtt ttttacacg ttacatacgt acctagtcg gaacgtaatt	24660
ttactaccgc tccagett tgcacgaag gtaaggcata ctttccacgc gaaggcgat	24720
tgcgttttaa cggtacatca tggtttatta cgcaacgtaa ttttttagt ccacaatta	24780
ttactactga taatacatc gttagcggtt attgcgtatg cgttacgtt attattaata	24840
ataccgtta ccatccatg caacctgagt tagactctt taaagaggaa ctcgataagt	24900
atttaaaaaa tcatacatca cctgacgtt acttagcga tatttcaggat attaagcctt	24960
cagtcgttaa tattcagaaa gagattgata gacttaacga agtcgtaaa aatcttaacg	25020
aatcacttat ccatctgca gatgttagta agtacgaaca gtatattaa tggccttgg	25080
acgtttgggtt agggtttata gccggctaa tgcgtatcgt tatggttacg atactgttat	25140
gttgtatgac atcatgttgt tcatgtctt aaggcgcattt ttcatgcggaa tcatgttgta	25200
aattcgcgca agacgattct gagccagtcg ttaaggaggat gaaattgcata tatacataaa	25260
cgaacttatg gatttgttta tgagatttt tactcttaga tcaattactg cacagccagt	25320
aaaaattgac aatgcttc tgcgaagttt acagcaacga taccgttaca	25380
agcctcactc ctttcggat ggcttgcattt tggcgttgc tttcttgc tttttcagag	25440
cgttaccaaa ataattgcgc tcaataaaag atggcageta gcccattata agggttcca	25500
gttcatttgc aatttactgc tgctattttt taccatctat tccatcttt tgcttgc	25560
tgcaggatg gaggcgcaat tttgtaccc tcatgcctt atatatttc tacaatgcata	25620
caacgcgtt aagaattttt tgagatgtt gcttggggaa aagtgcataat ccaagaaccc	25680
attactttat gatgcactt actttgtttt ctggcacaca cataactatg actactgtat	25740
accatataac aatgttcacag atacaattgt cgttactgaa ggtgacggca tttcaacacc	25800
aaaactcaaa gaagactacc aaattgggtt ttattctgat gataggactt cagggtttaa	25860
agactatgtc gttgtacatg gctatttc acgattttac taccagctt agtctacaca	25920
aattactaca gacactggta ttgaaatgc tacatttc acatcttac agcttggtaa	25980
agacccaccc aatgtgcataa tacacacaat cgacggctt tcaggagttt ctaatccagc	26040
aatggatcca atttatgtatg agccgacgac gactactgc gtgccttgc aagcacaaga	26100
aagtgtgtac gaacttatgtt actcatttcg ttcggaaagaa acaggatcgt taatagttaa	26160
tagcgtactt ctttttctt ctttcgttgcattt ttcgttgcata gtcacacttgc ccatcttac	26220
tgcgttgcgtt ttgtgtgcgtt actgtgttgc tattgttac gtcgttgcgtt taaaaccaac	26280
ggtttacgtc tactcgcgtt ttggggatctt gatgttgcgtt tttttttttt gatgtttttt	26340
ggtctaaacg aactaactat tattttttt ctgtttggaa cttaacattt gttatcatg	26400
ggcgataacg gtacaattac cggttggaa cttaacactt tactcgcata atggatctc	26460
gtatactgtt ttcgttgcgtt cgttgcgtt atgttgcata tttttttttt gatgtttttt	26520
aatcggtttt tgcgttgcgtt tttttttttt tttttttttt tttttttttt gatgtttttt	26580
gcgttgcgtt tactcgcgtt cgttgcgtt atgttgcata tttttttttt gatgtttttt	26640
gtatgggtt gtatgttcgg actttatgtgg ttgttctt ttcgttgcgtt atttatgtt	26700
ttcgacgtt tttttttttt tttttttttt tttttttttt tttttttttt gatgtttttt	26760

-continued

ccacttagag gtacaatcgt tactagacca cttatggagt ctgagttgt gataggcgca	26820
gtgattatta gggggcattt gcgtatggcc ggtcatagtc taggttagatg cgatattaag	26880
gatctaccta aagagattac cgttgcgact agtcgtacac tatcttattta taagttaggc	26940
gctagtcaac gtgtcggtac ttagatcgga ttgcagctt ataatacgta tcgtatcggt	27000
aattataagc ttaataccga tcacgctgga tctaaccgata atatcgattt actcgtacag	27060
taagtgacaa cagatgttc atcttgttga cttccaggtt acaatagcag agatattgtat	27120
tatcattatg aggactttca ggattgttat ttggaaatctt gacgttataa taagttcaat	27180
agttagacaa ttatthaagc ctcttaactaa gaagaattat tcggagtttag atgatgaaga	27240
acctatggag ttagattatc cataaaacga acatgaaaat tattctcttc ctgacatgt	27300
ttgttatttc atcttgcgag ctatatactt atcaggagtg tgtagaggt acgactgtac	27360
tactaaaaga accttgccc tcaggaacat acgagggcaa ttcaccatc caccctcttgc	27420
ctgacaataa atttgcacta acttgcacta gcacacactt tgctttgtc tgtgtgcacg	27480
gtactcgaca tacatcatcg ctgcgtgca gatcagtttc accaaaaactt ttcatcagac	27540
aagaggagggt tcaacaagag ctctactcgc cacttttttctt cattgttgc gctctagttat	27600
tttaataact ttgcttcacc attaagagaa agacagaatg aatgagctca cttaaatgt	27660
cttctattttgc tgcttttttag cctttctgtt attccttgc ttaataatgc ttattatatt	27720
ttgggtttca ctcgaaatcc aggtatctaga agaaccttgc accaaaagtctt aaacgaaat	27780
gaaaccttctc attgttttgc ctgttgcattt tctatgcagt tgcatatgc ctgtgtac	27840
gcgcgtgtca tctaataaac ctcatgtgt tgaagatctt tgtaggtac aacactaggg	27900
gtaataactta tagcactgct tggctttgtc ctcttagaaaa ggtttacact tttcatagat	27960
ggcacactat ggttcaaaaca tgcacaccta atgttactat caactgtcaa gatccagctg	28020
gtgggtgcgtc tatacgtagg tggttgtacc ttcatgaagg tcaccaaact gctgcattt	28080
gagacgtact tggtgttttta aataaacgaa caaattaaaaa tgtctgataa tggaccccaa	28140
tcaaaccac gtagtgccttcc ccgcattaca tttgggtggac ccacagattc aactgacaat	28200
aaccagaatg gaggacgc当地 tggggcaagg cccaaacagc gcccacccca aggtttaccc	28260
aataataactg cgtttgggtt cacagcttc actcagcatg gcaaggagga acttagattc	28320
cctcgaggcc agggcggtcc aatcaacacc aatagtggtc catatgcacca aattggctac	28380
taccgaagag ctacccgacg agttcggtt ggtgacggca aatgaaaga gctcagcccc	28440
agatggtaact tctattaccc aggaactggc ccagaagctt cacttccctt cggcgctaac	28500
aaagaaggca tcgtatgggt tgcaactgag ggagccttgc atacacccaa agaccacatt	28560
ggcacccgc当地 atcctaataa caatgtgtcc accgtgtac aacttccctca aggaacaaca	28620
ttgccaaaag gcttctacgc agagggaaagc agaggcggca gtcaagccctt ttctcgctcc	28680
tcatcacgtt gtcggtttaa ttcaagaaat tcaactcttc gcagcgttag gggaaattct	28740
cctgctcgaa tggctagcgg aggtggtaa actgcctcg cgctattgtc gctagacaga	28800
ttgaaccacg ttgagacaa agtttctgtt aaaggccaaac aacaacaagg cccaaactgtc	28860
actaagaaat ctgctgttgc ggcacatctaaa aagcctcgcc aaaaacgtac tgccacaaaa	28920
cagttacaacg tcactcaagc atttgggaga cgtggtccag aacaaacccca aggaaatttc	28980
ggggaccaag acctaatacg acaaggaact gattacaacat ttggccgc当地 aattgcacaa	29040
tttgctccaa gtgcctctgc attcttttgc atgtcacgca ttggcatgga agtcacaccc	29100
tcgggaacat ggctgactta tcatggagcc attaaattgg atgacaaaga tccacaattc	29160

-continued

```

aaagacaacg tcatactgct gaacaaggcac attgacgcatt cccaccaaca 29220
gaggcctaaaa aggacaaaaaa gaaaaagact gatgaagctc agccttgcgc gcagagacaa 29280
aagaaggcagc ccactgtgac tcttttcctt gcggttgaca tggatgattt ctccagacaa 29340
cttcggaaattt ccatggatgg agcttctgtt gattcaactc aggcataaac actcatgtt 29400
accacacaag gcagatgggc tatgtaaacg ttttcgcaat tccgtttacg atacatagtc 29460
tactcttgcg cagaatgttact tctcgtaact aaacagcaca agtaggttta gttaacttta 29520
atctcacata gcaatcttta atcaatgtt aacattaggg aggacttggaa agagccacca 29580
cattttcatc gagggcacgc ggagtacgtt cgagggtaca gtgaataatg ctaggggag 29640
ctgcctatat ggaagagccc taatgtgtaa aattaattttt agtagtgcta tccccatgtt 29700
attttaatag ttcttagga gaatgacaaa aaaaaaaaaa aaaaaaaaaa a 29751

```

20

We claim:

1. A modified virus which comprises a viral genome comprising a modified virus protein-encoding sequence having synonymous codons from the coding sequence of a parent virus in a rearranged order, wherein said rearrangement provides a reduced codon pair bias relative to a mammalian host over the coding sequence, in comparison to the coding region of the parent virus, without changing codon usage of the parent virus, wherein the codon pair bias is calculated by the following formula:

$$CPB = \sum_{i=1}^k \frac{CPS_i}{k-1}$$

wherein, the codon pair bias (CPB) of a protein encoding sequence is the arithmetic mean of the codon pair scores (CPS) of the individual codon pairs (i) contained within said protein encoding sequence of k codons in length; and wherein the modified protein encoding sequence has a codon pair bias at least 0.05 less than the codon pair bias of the parent protein encoding sequence.

2. The modified virus of claim 1, wherein the modified protein encoding sequence has a codon pair bias at least 0.1 less than the codon pair bias of the parent protein encoding sequence.

3. The modified virus of claim 1, wherein the modified protein encoding sequence has a codon pair bias at least 0.2 less than the codon pair bias of the parent protein encoding sequence.

4. The modified virus of claim 1, wherein the modified protein encoding sequence has a codon pair bias of -0.05 or less.

5. The modified virus of claim 1, wherein the modified protein encoding sequence has a codon pair bias of -0.1 or less.

6. The modified virus of claim 1, wherein the modified protein encoding sequence has a codon pair bias of -0.3 or less.

7. The modified virus of claim 1, wherein the modified protein encoding sequence has a codon pair bias of -0.4 or less.

8. The modified virus of claim 1, wherein the modified protein encoding sequence is less than 90% identical to the protein encoding sequence of the parent virus.

9. The modified virus of claim 1, wherein the modified protein encoding sequence is less than 80% identical to the protein encoding sequence of the parent virus.

10. The modified virus of claim 1, wherein the modified protein encoding sequence and the parent protein encoding sequence encode the same protein.

11. The modified virus of claim 1, wherein the modified protein encoding sequence encodes a protein that differs from a natural isolate by 10 amino acids or fewer.

12. The modified virus of claim 1, wherein the modified protein encoding sequence encodes a protein that differs from a natural isolate by 20 amino acids or fewer.

13. The modified virus of claim 1, wherein the modified protein encoding sequence is modified over a length of at least 100 nucleotides.

14. The modified virus of claim 1, wherein the modified protein encoding sequence is modified over a length of at least 500 nucleotides.

15. The modified virus of claim 1, wherein the modified protein encoding sequence is modified over a length of at least 1000 nucleotides.

16. The modified virus of claim 1, wherein the virus infects a human.

17. The modified virus of claim 1, wherein the virus induces an immune response in an animal host.

18. The modified virus of claim 1, wherein reversion of the modified protein encoding sequence to the wild-type sequence is inhibited when grown in a host.

19. The modified virus of claim 1, wherein the virus is a DNA, RNA, double-stranded, or single-stranded virus.

20. The modified virus of claim 1, wherein the protein encoding sequence encodes a polyprotein.

21. The modified virus of claim 1, wherein the nucleotide sequence encodes a capsid protein.

22. The modified virus of claim 1, wherein the virus is a poliovirus, rhinovirus, influenza virus, severe acute respiratory syndrome (SARS) coronavirus, Human Immunodeficiency Virus (HIV), Hepatitis C Virus (HCV), infectious bronchitis virus, Ebolavirus, Marburg virus, dengue fever virus, West Nile disease virus, Epstein-Barr virus (EBV), or yellow fever virus.

23. The modified virus of claim 1, wherein the virus is a Poxvirus, Herpes virus, Papillomavirus, or Adenovirus.

24. The modified virus of claim 22, wherein the poliovirus is derived from poliovirus type 1 (Mahoney), poliovirus type

281

2 (Lansing), poliovirus type 3 (Leon), monovalent oral poliovirus vaccine (OPV) virus, or trivalent OPV virus.

25. The modified virus of claim 1, wherein the codon pair rearrangements are distributed throughout the modified protein encoding sequence.

26. The modified virus of claim 1, wherein the codon pair rearrangements are restricted to a portion of the modified protein encoding sequence.

27. The modified virus of claim 25, wherein the protein encoding sequence encodes one or more capsid proteins.

28. The modified virus of claim 1, wherein the virus is a picornavirus.

29. The modified virus of claim 28, wherein the picornavirus is poliovirus.

30. The modified virus of claim 29, wherein the nucleotide sequence of the viral genome comprises SEQ ID NO:4 or a portion thereof.

31. The modified virus of claim 29, wherein the nucleotide sequence of the viral genome comprises SEQ ID NO:6.

32. The modified virus of claim 28, wherein the picornavirus is a rhinovirus.

33. The modified virus of claim 32, wherein the nucleotide sequence of the viral genome comprises SEQ ID NO:30 or SEQ ID NO:32 or a portion thereof.

34. The modified virus of claim 1, wherein the virus is an orthomyxovirus.

35. The modified virus of claim 34, wherein the orthomyxovirus is influenza virus.

36. The modified virus of claim 35, wherein the nucleotide sequence of the viral genome comprises one or more of SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:28, and portions thereof.

37. The modified virus of claim 35, wherein the nucleotide sequence of the viral genome comprises SEQ ID NO:23.

38. The modified virus of claim 35, wherein the nucleotide sequence of the viral genome comprises SEQ ID NO:15 and SEQ ID NO:21.

39. The modified virus of claim 1, wherein the virus is a flavivirus.

40. The modified virus of claim 39, wherein the flavivirus is Dengue virus.

41. The modified virus of claim 40, wherein the nucleotide sequence of the viral genome comprises SEQ ID NO:34.

42. The modified virus of claim 1, wherein the virus is a retrovirus.

43. The modified virus of claim 42, wherein the retrovirus is human immunodeficiency virus (HIV).

282

44. The modified virus of claim 43, wherein the nucleotide sequence of the viral genome comprises SEQ ID NO:36.

45. The modified virus of claim 1, wherein the virus is a reovirus.

46. The modified virus of claim 45, wherein the reovirus is rotavirus.

47. The modified virus of claim 46, wherein the nucleotide sequence of the viral genome comprises SEQ ID NO:38 and/or SEQ ID NO:40.

48. The modified virus of claim 1, wherein the virus is a coronavirus.

49. The modified virus of claim 48, wherein the coronavirus is severe acute respiratory syndrome (SARS) virus.

50. The modified virus of claim 49, wherein the nucleotide sequence of the viral genome comprises SEQ ID NO:42.

51. An immunogenic composition for inducing an immune response in a subject comprising the modified virus of claim 1, and a pharmaceutically acceptable carrier.

52. The composition of claim 51, wherein the modified virus exhibits lower virion-specific infectivity than the wild type virus.

53. The composition of claim 51, wherein the modified virus induces the same immune response in a host animal as the corresponding wild type virus.

54. The modified virus of claim 1, wherein the modified protein encoding sequence has a codon pair bias at least 0.3 less than the codon pair bias of the parent protein encoding sequence.

55. The modified virus of claim 1, wherein the modified protein encoding sequence has a codon pair bias at least 0.4 less than the codon pair bias of the parent protein encoding sequence.

56. The modified virus of claim 1, wherein codon pair bias is reduced by rearranging existing codons of the parent protein-encoding sequence.

57. The modified virus of claim 1, wherein the rearranged codons provide a codon pair bias reduction of at least 0.05.

58. The modified virus of claim 1, wherein the rearranged codons provide a codon pair bias reduction of at least 0.1.

59. The modified virus of claim 1, wherein the rearranged codons provide a codon pair bias reduction of at least 0.2.

60. The modified virus of claim 1, wherein the rearranged codons provide a codon pair bias reduction of at least 0.3.

61. The modified virus of claim 1, wherein the rearranged codons provide a codon pair bias reduction of at least 0.4.

62. The modified virus of claim 1, wherein the codon usage over the protein coding region remains unchanged.

* * * * *